

**O'Bryen, Barbara**

---

**From:** Switzer, Juliet  
**Sent:** Thursday, December 18, 2003 8:37 AM  
**To:** O'Bryen, Barbara  
**Subject:** please search

09/889491

please search seq id no 13 and 14 in all prior art databases. please also search in reg file in oligos of 100 or fewer crossed as appropriate with alignments. Thanks.

Juliet Switzer  
Art Unit 1634  
703 306 5824  
office CM1 12D15  
mailbox CM1 12E12

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: December 18, 2003, 09:02:17 / Search time 1387.89 Seconds

(without alignments)  
913.763 Million cell updates/sec

Title: US-09-889-491-14

Perfect score: 31

Sequence: 1 atagtgaaactcgtgtaattatgaatttt 31

Scoring table:

IDENTITY NUC

Gapop 10.0, Gapept 1.0

Searched: 2888711 seqs, 2045481386 residues

5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database:

GenBank1:

1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_ov:\*  
22: em\_or:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hg\_hum:\*  
31: em\_hg\_in:\*  
32: em\_hg\_oth:\*  
33: em\_hg\_mus:\*  
34: em\_hg\_pln:\*  
35: em\_hg\_rod:\*  
36: em\_hg\_mam:\*  
37: em\_hg\_vit:\*  
38: em\_sy:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_oth:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	100.0	31	6	AX028511
2	31	100.0	2415	9	HUMNSPOL
3	29.4	94.8	1221	9	HUMSIALPOL
4	29.4	94.8	2461	2	AP49945885
5	29.4	94.8	78036	2	AC090084
6	29.4	94.8	187624	9	AC093768
7	27.8	89.7	7771	6	AX346874
8	23	74.2	153586	2	AC046875
9	23	74.2	155229	2	AC016316
10	23	74.2	165217	2	AC116813
11	23	74.2	168433	9	AC132362
12	23	74.2	173519	2	AC104741
13	23	74.2	181191	2	AC080138
14	23	74.2	202233	9	AC093864
15	23	74.2	216280	2	AC098864
16	23	74.2	217879	2	EX005264
17	23	74.2	230632	2	AC096365
18	23	74.2	293264	1	AB017031
19	23	72.9	106245	2	AC139706
20	22.6	72.9	114264	2	AC141436
21	22.6	72.9	172434	8	AC090486
22	22.6	72.9	302524	8	AB017075
23	22.2	71.6	42861	3	CEP52B11
24	22.2	71.6	110000	2	AC125548
25	22.2	71.6	253819	2	AC095474
26	22.2	71.6	310193	2	AC131872
27	22.2	71.6	8694	14	AF538602
28	22	71.0	89004	3	CEY69H2
29	22	71.0	110000	3	AC116984
30	22	71.0	140765	9	AC021301
31	22	71.0	167749	2	AC021489
32	22	71.0	177503	9	AC005297
33	22	71.0	186114	9	AC021621
34	22	71.0	190359	2	AC118356
35	22	71.0	192407	5	AL928827
36	22	71.0	194398	10	AL627075
37	22	71.0	207495	10	AL772393
38	22	71.0	222699	2	AC129158
39	22	71.0	231284	2	AC109415
40	22	71.0	234004	2	AC115778
41	22	71.0	238344	2	AC111935
42	22	71.0	263608	2	AL772254
43	21.8	70.3	282183	1	CU11168X4
44	21.6	69.7	618	11	BV053839
45					

## ALIGNMENTS

RESULT 1	AX028511	31 bp	DNA	linear	PAT 16-SEP-2000
LOCUS	AX028511				
DEFINITION	Sequence 14 from Patent WO0042216.				
ACCESSION	AX028511				
VERSION	AX028511.1				
KEYWORDS	GI:10189694				
SOURCE					
ORGANISM	Homo sapiens (human)				
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	Kueh, P.				
JOURNAL	Genetic predisposition				
	Patent: WO 0042216-A 14 20-JUL-2000;				

FEATURES  
source

OSTEOMETER BIOTECH AS (DK) ; KUSK PHILIP (DK)  
Location/Qualifiers  
1..31  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 12 a 1 c 5 g 13 t

Query Match  
Best Local Similarity 100.0%; Score 31; DB 6; Length 31;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAGTGAACCTTGTGTATTTATGAAATTTT 31  
1 ATAGTGAACCTTGTGTATTTATGAAATTTT 31

RESULT 2

HUMNSP01 2415 bp DNA linear PRI 29-SEP-1994  
LOCUS Human bone sialoprotein (BNSP) gene, exon 1.  
DEFINITION L24756.1 GI:438617  
ACCESSION L24756.1 GI:438617  
VERSION L24756.1 GI:438617  
KEYWORDS sialoprotein.  
SEGMENT 1 of 4  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2415)  
AUTHORS Kim, R.H., Shapiro, H.S., Li, J.J., Wzana, J.L. and Sodek, J.  
TITLE Characterization of the human bone sialoprotein (BSP) gene and its promoter sequence  
JOURNAL Matrix Biol. 14 (1), 31-40 (1994)  
MEDLINE 94340202  
PubMed 8061918

COMMENT Original source text: Homo sapiens DNA.  
FEATURES Location/Qualifiers

source

1..2415  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/map="Unassigned"  
2179..2260  
/gene="BNSP"  
/number=1  
exon

BASE COUNT 890 a 437 c 423 g 665 t

ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 31; DB 9; Length 2415;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAGTGAACCTTGTGTATTTATGAAATTTT 31  
1854 ATAGTGAACCTTGTGTATTTATGAAATTTT 1884

RESULT 3

HUMSIALP01 1221 bp DNA linear PRI 13-JAN-1995  
LOCUS Homo sapiens bone sialoprotein (BNSP) gene, exon 1.  
DEFINITION L09554.1 GI:292488  
ACCESSION L09554.1 GI:292488  
VERSION L09554.1 GI:292488  
KEYWORDS sialoprotein.  
SEGMENT 1 of 5  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1221)  
AUTHORS Kerr, J.M., Fisher, L.W., Termini, J.D., Wang, M.G., McBride, O.W. and

TITLE Young, M.F.  
The human bone sialoprotein gene (IBSP): genomic localization and characterization

JOURNAL Genomics 17 (2), 408-415 (1993)  
MEDLINE 94010937  
PubMed 8406493

COMMENT Original source text: Homo sapiens lung DNA.  
FEATURES Location/Qualifiers

source

1..1221  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/map="4q11-q21"  
/cell\_line="WI38"  
/cell\_type="fibroblast"  
/tissue\_type="lung"  
344..427  
/gene="IBSP"  
/note="G00-118-889"  
/number=1

exon

BASE COUNT 376 a 206 c 185 g 454 t

ORIGIN

Query Match  
Best Local Similarity 94.8%; Score 29.4; DB 9; Length 1221;  
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATAGTGAACCTTGTGTATTTATGAAATTTT 31  
21 ATAGTGAACCTTGTGTATTTATGAAATTTT 51

RESULT 4

AF4994583 2451 bp DNA linear PRI 02-DEC-2002  
LOCUS Homo sapiens bone sialoprotein (IBSP) gene, exon 1.  
DEFINITION AF499460  
ACCESSION AF499460  
VERSION AF499460.1 GI:20453970  
KEYWORDS  
SEGMENT 3 of 3  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2451)  
AUTHORS Kiyoshima, T., Yamauchi, M., Wong, C., Theon, A., Ganss, B. and Sodek, J.  
TITLE An  $\alpha$  element disrupts human bone sialoprotein promoter: lack of tissue-specific regulation by distalless (Dlx) and runt homeodomain protein2 (Runx2)/core binding factor  $\alpha$ 1 (Cbfa1) elements

JOURNAL Gene 299 (1-2), 205-217 (2002)  
PubMed 12459268

REFERENCE 2 (bases 1 to 2461)  
AUTHORS Kiyoshima, T. and Sodek, J.  
TITLE Direct Substitution

Submitted (08-APR-2002) CIHR Group in Matrix Dynamics, University of Toronto, Room 234, 150 College St, Toronto, ON M5S3E2, Canada  
FEATURES Location/Qualifiers

source

1..2461  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4q21.2-q21.3"  
/note="BAC NH0118108"  
/note="clone provided by CIHR Genome Facility, Hospital for Sick Children, Toronto, ON, Canada"

gene

CANT\_signal

TATA\_signal

/gene="IBSP"  
 /note="Inverted"  
 2106. >2187  
 /gene="IBSP"  
 /product="bone sialoprotein"  
 2106. 2187  
 /gene="IBSP"  
 /number=1  
 2188. >2461  
 /gene="IBSP"  
 /number=1  
 BASE COUNT 891 a 447 c 426 g 691 t 6 others  
 ORIGIN  
 Query Match 94.8%; Score 29.4; DB 9; Length 2461;  
 Best Local Similarity 96.8%; Pred. No. 5.3;  
 Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATAGTGAAGAACTGTGTATTAATTAATTT 31  
 1781 ATAGTGAAGAACTGTGTATTAATTAATTT 1811  
 RESULT 5  
 LOCUS AC090084 78036 bp DNA linear HTG 14-FEB-2001  
 DEFINITION Homo sapiens chromosome 4 clone CTD-2363K13 map 4, WORKING DRAFT  
 AC090084  
 AC090084.1 GI:12830225  
 HTG; HTGS PHASE1; HTGS\_DRAFT.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1. (bases 1 to 78036)  
 Bitten, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 4, clone CTD-2363K13  
 Unpublished  
 2 (bases 1 to 78036)  
 Bitten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
 Barina, N., Baetli, V., Boguslavskiy, L., Bouckhelter, B., Brown, A.,  
 Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collamore, A., Cooke, P., Dekrellano, K., Dewar, K., Diaz, J. S.,  
 Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,  
 Gardyna, S., Glend, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hedges, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Karatas, A., LaRocque, K., Lamas, R., Landers, T.,  
 Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,  
 Marquis, N., Matthews, C., McCarthy, M., McGowan, P., McKernan, K.,  
 McHeaters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunhahang, P., Pletre, N., Pollara, V., Raymond, C., Retta, R.,  
 Ribback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,  
 Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P.,  
 Sougnier, C., Spencer, B., Strange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talamas, J., Teafey, S., Theodore, J.,  
 Travers, M., Travis, N., Triggillo, U., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (14-FEB-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 Center: Genome Center  
 Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information

TITLE  
 JOURNAL  
 COMMENT

Center project name: L10982  
 Center clone name: 2363\_K13  
 Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 7762 bases at least Q40  
 Consensus quality: 7767 bases at least Q30  
 Consensus quality: 77801 bases at least Q20  
 Insert size: 7900; agarose-fp  
 Insert size: 77836; sum-of-ctnigs  
 Quality coverage: 8.0 in Q20 bases; agarose-fp  
 Quality coverage: 8.1 in Q20 bases; sum-of-ctnigs  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 2322: contig of 2322 bp in length  
 \* 2323 2422: gap of 100 bp  
 \* 2423 12454: contig of 10042 bp in length  
 \* 12465 12564: gap of 100 bp  
 \* 12565 78036: contig of 65472 bp in length.  
 Location/Qualifiers  
 1. 78036  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /map="4"  
 /chromosome="4"  
 /clone="CTD-2363K13"  
 /clone\_1ib="CTD Human BAC"  
 1. 2322  
 /note="assembly\_fragment"  
 2423. 12464  
 /note="assembly\_fragment"  
 12565. 78036  
 /note="assembly\_fragment"  
 BASE COUNT 22977 a 14696 c 14398 g 25765 t 200 others  
 ORIGIN

misc\_feature  
 misc\_feature  
 misc\_feature  
 misc\_feature  
 Query Match 94.8%; Score 29.4; DB 2; Length 78036;  
 Best Local Similarity 96.8%; Pred. No. 2.1;  
 Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATAGTGAAGAACTGTGTATTAATTAATTT 31  
 18081 ATAGTGAAGAACTGTGTATTAATTAATTT 18051  
 RESULT 6  
 LOCUS AC093768 187624 bp DNA linear PRI 01-MAR-2002  
 DEFINITION Homo sapiens BAC clone RP11-115G13 from 4, complete sequence.  
 AC093768 AC021959  
 AC093768.3 GI:16973767  
 HTG.  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 187624)  
 Sulston, J. E., and Waterston, R.  
 Toward a complete human genome sequence  
 Genome Res. 8 (11), 1097-1108 (1998)  
 98063792  
 MEDLINE  
 PUBMED  
 9847074  
 2 (bases 1 to 187624)  
 Harkins, C., Kozlowski, A. and Doebber, A.

TITLE	The sequence of Homo sapiens BAC clone R11-113G13
JOURNAL	Unpublished (2001)
REFERENCE	3 (bases 1 to 187624)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (10-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA
REFERENCE	4 (bases 1 to 187624)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (11-NOV-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	5 (bases 1 to 187624)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	6 (bases 1 to 187624)
AUTHORS	Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT	On Nov 17, 2001 this sequence version replaced gi:164183232.

NOTES: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:** The RPEC-11 human BAC library was made from the blood of one male donor, as described by Oseegawa, K., Moon, P. Y., Zhao, B., Frangen, E., Tatemoto, M., Catanesi, U. J. and de Jong, P. O. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is AC087106. Actual start of this clone is at base position 1 of Rp11-113G13; actual end is at base position 187624 of Rp11-113G13.

Sequence derived from one plasmid subclone, base position 69516 to 69532.

**FEATURES**  
**SOURCE**

The sequence of AC021959 has been incorporated into AC093768.

source	Location/Qualifiers
repeat_region	1. 187624
repeat_region	/organism="Homo sapiens"
repeat_region	/mol_type="genomic DNA"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="4"
repeat_region	/map="4"
repeat_region	/clone="RP11-113G13"
repeat_region	/clone_1lb="RPCT-11"
repeat_region	6069. 6195
repeat_region	/rpt_family="L1"
repeat_region	6288. 6308
repeat_region	/rpt_family="AT_rich"
repeat_region	6422. 6530
repeat_region	/rpt_family="L1"
repeat_region	6559. 6582
repeat_region	/rpt_family="AT_rich"
repeat_region	6708. 6744
repeat_region	/rpt_family="AT_rich"
repeat_region	6732. 7028
repeat_region	/rpt_family="Alu"
repeat_region	8033. 8053
repeat_region	/rpt_family="TTTTAA)n"
repeat_region	8041. 8459
repeat_region	/rpt_family="L1"
repeat_region	8804. 8942
repeat_region	/rpt_family="MIR"
repeat_region	8943. 9214
repeat_region	/rpt_family="MAIR"
repeat_region	9526. 9591
repeat_region	/rpt_family="(TA)n"
repeat_region	10055. 10530
repeat_region	/rpt_family="L1"
repeat_region	11326. 11417
repeat_region	/rpt_family="L2"
repeat_region	11399. 11445
repeat_region	/rpt_family="(TGA)n"
repeat_region	15645. 15865
repeat_region	/rpt_family="L2"
repeat_region	15933. 16227
repeat_region	/rpt_family="Alu"
repeat_region	complement(1610. 16161)
repeat_region	/note="similar to Homo sapiens EST A1745652 (NID:G5113940)
repeat_region	tr24c10.x1"
repeat_region	16226. 16281
repeat_region	/rpt_family="(CAA)n"
repeat_region	16337. 16436
repeat_region	/note="similar to Homo sapiens EST AA873157 (NID:G2965279)
repeat_region	0173g11.61"
repeat_region	16340. 16915
repeat_region	/note="similar to Homo sapiens EST AV702108 (NID:G10718438)"
repeat_region	16340. 16899
repeat_region	/note="similar to Sus scrofa EST BT181092 (NID:G14655501)"
repeat_region	16340. 16447
repeat_region	/note="similar to Rattus norvegicus EST BF391035 (NID:G11375882)"
repeat_region	16340. 16436
repeat_region	/note="similar to Homo sapiens EST AA651669 (NID:G25833221)
repeat_region	na89604.61"
repeat_region	16340. 16436
repeat_region	/note="similar to Homo sapiens EST AA885538 (NID:G30006466)
repeat_region	nz5202.61"
repeat_region	16340. 16436
repeat_region	/note="similar to Homo sapiens EST A1088133 (NID:G3427109)
repeat_region	0295g07.x1"
repeat_region	16340. 16436
repeat_region	/note="similar to Homo sapiens EST AL533342 (NID:G12796735)"
repeat_region	16340. 16436
repeat_region	/note="similar to Homo sapiens EST BE742618 (NID:G10156597)"

```

misc_feature 16340..16436
/note="similar to Homo sapiens EST B1094057
(NID:g14512387)"
misc_feature 16342..16436
/note="similar to Homo sapiens EST AA398444 (NID:g2051756)
zt74907.s1"
misc_feature 16355..16380
/note="similar to Homo sapiens EST W00667 (NID:g1272102)
za72b10.x1"
misc_feature 16388..16436
/note="similar to Homo sapiens EST BE621830
(NID:g9892770)"
misc_feature 16403..17080
/note="similar to Homo sapiens EST B1084966
(NID:g14503296)"
misc_feature 16405..16436
/note="similar to Homo sapiens EST B1522324
(NID:g15347116)"
misc_feature 16406..16436
/note="similar to Homo sapiens EST A1745652 (NID:g5113940)
tr24c10.x1"
misc_feature 16458..17080
/note="similar to Homo sapiens EST AU533242
(NID:g12796735)"
misc_feature 16458..17080
/note="similar to Homo sapiens EST BE621830
(NID:g9892770)"
misc_feature 16458..17080
/note="similar to Homo sapiens EST BE742618
(NID:g10156597)"
misc_feature 16458..17077
/note="similar to Homo sapiens EST B1522324
(NID:g15347116)"
misc_feature 16458..16873
/note="similar to Homo sapiens EST AA398444 (NID:g2051756)
zt74907.s1"
misc_feature 16458..16868
Query Match 94.8%; Score 29.4; DB 9; Length 187624;
Best Local Similarity 96.8%; Pred. No. 1.7;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATAGTGAACCTGTGTAATTAATTAATTTT 31
Db 29926 ATAGTGAACCTGTGTAATTAATTAATTTT 29956

RESULT 7
AX346874 7771 bp DNA linear PAT 01-FEB-2002
LOCUS AX346874
DEFINITION Sequence 1945 from Patent WO0200928.
ACCESSION AX346874
VERSION AX346874.1 GI:18494760
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS
Olek,A., Piepenbrock,C. and Berlin,K.
TITL
Diagnosis of diseases associated with the immune system
JOURNAL
Patent: WO 0200928-A 1945 03-JAN-2002;
Epidemiology AG (DB)
FEATURES
SOURCE
1..7771
location/Qualifiers
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2788 a 45 c 1303 g 3635 t
ORIGIN
Query Match 89.7%; Score 27.8; DB 6; Length 7771;
Best Local Similarity 93.5%; Pred. No. 13;

```

```

Matches 29; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATAGTGAACCTGTGTAATTAATTAATTTT 31
Db 4650 ATAGTGAACCTGTGTAATTAATTAATTTT 4680

RESULT 8
AX346875/c 7771 bp DNA linear PAT 01-FEB-2002
LOCUS AX346875
DEFINITION Sequence 1946 from Patent WO0200928.
ACCESSION AX346875
VERSION AX346875.1 GI:18494761
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS
Olek,A., Piepenbrock,C. and Berlin,K.
TITL
Diagnosis of diseases associated with the immune system
JOURNAL
Patent: WO 0200928-A 1946 03-JAN-2002;
Epidemiology AG (DB)
FEATURES
SOURCE
1..7771
location/Qualifiers
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2143 a 45 c 1537 g 4046 t
ORIGIN
Query Match 74.2%; Score 23; DB 6; Length 7771;
Best Local Similarity 83.9%; Pred. No. 4.7e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ATAGTGAACCTGTGTAATTAATTAATTTT 31
Db 3122 ATAGTGAACCTGTGTAATTAATTAATTTT 3092

RESULT 9
AC016316/c 153586 bp DNA linear HTG 09-SEP-2000
LOCUS AC016316
DEFINITION Homo sapiens clone RP11-28K18, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
ACCESSION AC016316
VERSION AC016316.3 GI:10045694
KEYWORDS
HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Bakayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 153586)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITL
Homo sapiens, clone RP11-28K18
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 153586)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,N.,
Baldwin,J., Batra,N., Beckwith,R., Boguslavsky,I., Bouckgalter,B.,
Brown,A., Castle,A., Collange,L., Collins,S., Collymore,A.,
Cooke,P., Dearrellano,K., Dewar,K., Donlin,M., Donnellan,L., Doyle,M.,
Ferreira,P., Fitzhugh,M., Forrest,C., Funke,R., Gage,D.,
Gallagan,J., Gardyna,S., Grant,G., Hages,B., Heathford,A., Horton,L.,
Howland,J.C., Johnson,K., Jones,C., Kann,J., Karatsis,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., MacDonald,P., Marquis,N.,
McSwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,D., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Sanger-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tittell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITL
Direct Submision
JOURNAL
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome

```

## COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 9, 2000 this sequence version replaced gi:6649327.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: 14776

Center clone name: 28 K.18

## Summary Statistics

Sequencing vector: M13, M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 149540 bases at least Q40  
 Consensus quality: 151261 bases at least Q30  
 Consensus quality: 151990 bases at least Q20  
 Insert size: 152000; agarose-fp  
 Insert size: 152786; sum-of-coverage  
 Quality coverage: 5.8 in Q20 bases; sum-of-coverage  
 Quality coverage: 5.8 in Q20 bases; sum-of-coverage

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 224: contig of 224 bp in length  
 \* 225 324: gap of 100 bp  
 \* 325 6074: contig of 5750 bp in length  
 \* 6075 6174: gap of 100 bp  
 \* 6175 37911: contig of 31737 bp in length  
 \* 37912 38011: gap of 100 bp  
 \* 38012 48659: contig of 10648 bp in length  
 \* 48660 48759: gap of 100 bp  
 \* 48760 70090: contig of 21331 bp in length  
 \* 70091 70190: gap of 100 bp  
 \* 70191 93247: contig of 23057 bp in length  
 \* 93248 93348: gap of 100 bp  
 \* 93349 118460: contig of 25113 bp in length  
 \* 118461 118560: gap of 100 bp  
 \* 118561 143777: contig of 25217 bp in length  
 \* 143778 143878: gap of 100 bp  
 \* 143879 153586: contig of 9709 bp in length.

## FEATURES

## source

1..153586

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/clone="RP11-28K18"

/clone\_lib="RP11-28K18 Human Male BAC"

1..224

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left"

misc\_feature

325..6074

/note="assembly\_fragment"

6175..37911

/note="assembly\_fragment"

38012..48659

/note="assembly\_fragment"

48760..70090

/note="assembly\_fragment"

70191..93247

/note="assembly\_fragment"

93348..118460

/note="assembly\_fragment"

misc\_feature 118561..143777  
 misc\_feature 143878..153586

/note="assembly\_fragment"  
 /note="assembly\_fragment"

clone\_end:T7

vector\_side:right"

BASE COUNT 46298 a 29228 c 29238 g 48021 t 801 others

## ORIGIN

Query Match 74.2%; Score 23; DB 2; length 153586;  
 Best Local Similarity 83.9%; Pred. No. 2.1e+02;  
 Matches 26; Conservatve 0; Mismatches 5; Indels 0; Gaps 0;

1 ATGTGAAACTGTGTGTAATTTGAAATTTT 31

Db 109288 ATGTGAAAGCTGTGTGAAATTTGAAATTTT 109258

## RESULT 10

## AC116813/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

## REFERENCE

## AUTHORS

## JOURNAL

## TITILE

Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Macleod, C., Macdonald, P., Major, J., Manning, D., Matthews, C., McCarthy, M., Melidim, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillo, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155229)  
 1. Birren, B., Nussbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArliano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J

Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Rachunka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zahoun, U., Zembek, L., Zimmer, A. and Zody, M.

Submitted (04-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 4, 2003 this sequence version replaced gi:21735088.  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WTR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: 125561

Center clone name: 357 G.23

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 154720 bases at least Q40

Consensus quality: 154857 bases at least Q30

Consensus quality: 154917 bases at least Q20

Insert size: 15000; agarose-fp

Insert size: 154929; sum-of-coverage

Quality coverage: 11.5 in Q20 bases; agarose-fp

Quality coverage: 11.1 in Q20 bases; sum-of-coverage

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given. However the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 30191: contig of 30191 bp in length  
\* 30192 30291: gap of 100 bp  
\* 30292 64548: contig of 34257 bp in length  
\* 64549 64648: gap of 100 bp  
\* 64649 124983: contig of 60335 bp in length  
\* 124984 125084: gap of 100 bp  
\* 125084 155229: contig of 30146 bp in length.

## FEATURES

## source

1. 155229

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="RP24-357G23"

/clone\_lib="RP24 Male Mouse BAC"

1. 30191

/note="assembly\_fragment"

clone end:3P6

vector\_side:left"

30292..64548

/note="assembly\_fragment"

64649..124983

/note="assembly\_fragment"

125084..155229

/note="assembly\_fragment"

clone end:17

vector\_side:right"

BASE COUNT 49696 a 30207 c 29719 g 45307 t 300 others

## ORIGIN

Query Match 74.2%; Score 23; DB 2; Length 155229;  
Best Local Similarity 83.9%; Pred. No. 21e+02;  
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ATATGTAACACTTGTGTATATATGAAATTTT 31

||||| 9674 ATATGTAACACTTGTGTATATATGAAATTTT 9644

## RESULT 11

AC132362

LOCUS

DEFINITION

AC132362

SEQUENCE, 2 unordered pieces.

AC132362

AC132362.2 GI:31560447

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ACTIVEFIN.

KEYWORDS

SOURCE

Mus musculus (house mouse)

ORGANISM

REFERENCE

WILSON, R.K.

1 (bases 1 to 165217)

2 (bases 1 to 165217)

3 (bases 1 to 165217)

4 (bases 1 to 165217)

5 (bases 1 to 165217)

6 (bases 1 to 165217)

7 (bases 1 to 165217)

8 (bases 1 to 165217)

9 (bases 1 to 165217)

10 (bases 1 to 165217)

11 (bases 1 to 165217)

12 (bases 1 to 165217)

13 (bases 1 to 165217)

14 (bases 1 to 165217)

15 (bases 1 to 165217)

16 (bases 1 to 165217)

17 (bases 1 to 165217)

18 (bases 1 to 165217)

19 (bases 1 to 165217)

20 (bases 1 to 165217)

21 (bases 1 to 165217)

22 (bases 1 to 165217)

23 (bases 1 to 165217)

24 (bases 1 to 165217)

25 (bases 1 to 165217)

26 (bases 1 to 165217)

27 (bases 1 to 165217)

28 (bases 1 to 165217)

29 (bases 1 to 165217)

30 (bases 1 to 165217)

31 (bases 1 to 165217)

32 (bases 1 to 165217)

33 (bases 1 to 165217)

34 (bases 1 to 165217)

35 (bases 1 to 165217)

36 (bases 1 to 165217)

37 (bases 1 to 165217)

38 (bases 1 to 165217)

39 (bases 1 to 165217)

40 (bases 1 to 165217)

41 (bases 1 to 165217)

42 (bases 1 to 165217)

43 (bases 1 to 165217)

44 (bases 1 to 165217)

45 (bases 1 to 165217)

## COMMENT

Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu

Contact: submissions@wustl.edu

Project Information

Center project name: M\_BA0455805

Summary Statistics

Sequencing vector: M13; 0%

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 164699 bases at least Q40

Consensus quality: 164713 bases at least Q30

Consensus quality: 164718 bases at least Q20

Insert size: 166000; agarose-fp

Insert size: 165117; sum-of-coverage

Quality coverage: 14.29 in Q20 bases; agarose-fp

Quality coverage: 13.37 in Q20 bases; sum-of-coverage

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 43860: contig of 43860 bp in length  
\* 43861 43960: gap of unknown length  
\* 43961 165217: contig of 121257 bp in length.

## FEATURES

## source

1. 165217

/organism="Mus musculus"



```

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-455B5"
misc_feature
1..43860
/feature="assembly_name:Contig6"
misc_feature
43961..165217
/feature="assembly_name:Contig7"
BASE COUNT
45810 a 35165 c 34056 g 50086 t 100 others
ORIGIN

```

```

Query Match
Best Local Similarity 83.9%; Pred. No. 2.1e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 ATAGTGAACCTTGTGTAATTAATGAAATTTT 31
DB 67609 ATGTGAAACCTTGTGTCATCTCTAAATTTT 67639

```

```

RESULT 12
LOCUS AC104741 168433 bp DNA linear PRI 28-MAR-2002
DEFINITION Homo sapiens chromosome 3 clone RP11-228H5, complete sequence.
ACCESSION AC104741 AC025537
VERSION AC104741.2 GI:19774518
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saepthummachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
Direct Submision
Unpublished
2 (bases 1 to 168433)
Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
Direct Submision
Submitted (21-DEC-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 168433)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saepthummachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
Direct Submision
Submitted (28-MAR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Mar 28, 2002 this sequence version replaced gi:17975270.
COMMENT
----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgctgs@u.washington.edu
Drafting Center: WUGSC
----- Project Information
Center project name: chr-3
Center clone name: RP11-228H5 (bc0314)
----- Summary Statistics
Sequencing vector: plasmid; 25% of reads
Sequencing vector: M13; 25% of reads
Sequencing vector: plasmid; 108752; 49% of reads
Chemistry: Dye-terminator BT; 26% of reads
Chemistry: Dye-terminator BT; 25% of reads
Chemistry: Dye-terminator Big Dye; 48% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 168386 bases at least Q40
Consensus quality: 168421 bases at least Q30
Consensus quality: 168433 bases at least Q20
Insert size: 168433; sum-of-contigs
Quality coverage: 12.2x in Q20 bases; sum-of-contigs
-----
Overlapping Sequences:
5'; Mapping in progress

```

3': RP11-14301 (UWGC:bc0262) AC025817

#### Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

#### Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector. In order to accurately represent the entire circular BAC, small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

SeqDerMap	Fingerprint	SeqDerMap	Fingerprint	SeqDerMap	Fingerprint
8696	8751	2419	2464	5987	6208
6	<800	6382	6539	2067	2097
1117	1122	512	<800	5089	5190
355	<800	449	<800	639	<800
818	838	2127	2104	4326	4394
1852	1884	1009	1031	5081	5190
2865	2911	6183	6195	4480	4574
6260	6352	2988	3042	3685	3710
9196	9141	2247	2308	2226	2298
346	<800	223	<800	8407	8494
3111	3138	14916	15066	7	<800
8365	8231	3658	3623	1192	1202
7817	7818	4175	4079	458	<800
6417	6352	725	<800	1215	1202
2649	2636	2321	2308	6163	6208
5174	5118	1194	1214	3491	3585
3260	3283	2650	2660	699	<800
5737	5748	542	<800	272	<800

2361	2361	3101	3042	10526	10473
897	931	4156	4079	708	<800
3865	3915	5463	5411	11660	11718
614	<800	1461	1442	653	<800
166	<800	1314	1340	5422	5553
662	<800	462	<800	1429	1403
2347	2361	820	760	1231	1202
3192	3138	9121	9090	13263	13297
3925	3915	3637	3623	5646	5803
4069	4047	851	858	783	<800
14779	14846	726	<800	1315	1276
5328	5118	1439	1442	3401	3370
5038	5118	2798	2761	5066	5190
3043	3138	796	<800	80	<800
3124	3138	5195	5177	498	<800
4499	4489	1227	1214	3337	3370
785	<800	2270	2308	748	<800
118	<800	1923	1900	1627	1612
403	<800	1368	1340	935	958
1866	1884	2604	2660	5425	5553
8184	8231	138	<800	1713	1727
3178	3138	3645	3623	2219	2298
1483	1479	833	858	3058	3155
7092	7039	604	<800	788	<800
1113	1122	7412	7327	2756	2835
672	<800	1885	1900	1292	1276
8776	8751	657	<800	4807	5190
4512	4489	1808	1900	9186	9251
7000	7039	12834	12694	426	<800
		1402	1442	559	<800
		853	858	3605	3585
		2705	2660	436	<800
		1718	1703	328	<800
		165	<800	10607	10473
		100	<800	1126	1110
		2639	2660	4989	5190
		3005	3042		

Query Match 74.2%; Score 23; DB 9; Length 168433;  
 Best Local Similarity 83.9%; Pred. No. 2, 1e+02;  
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGAGAAACCTTGCTGTAATTAAGAAATTTT 31  
 Db 155208 ACAGTGACACCTTTATTAATTAAGAAATTTT 155178

RESULT 13  
 AC130840 173519 bp DNA linear HTG 14-AUG-2002  
 AC130840  
 LOCUS Mus musculus chromosome UNK clone RP24-443G4, WORKING DRAFT  
 DEFINITION SEQUENCE, 21 unordered pieces.  
 AC130840  
 VERSION AC130840.1 GI:22218609  
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 173519)  
 McPherson, J.D. and Waterston, R.H.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 173519)  
 McPherson, J.D. and Waterston, R.H.  
 DIRECT SUBMISSION  
 TITLES Direct Submission  
 JOURNAL Submitted (14-AUG-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA

## COMMENT

Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc/index.shtml>  
 Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
 Project Information  
 Center project name: M BB0443G04

Summary Statistics  
 Sequencing vector: MJ3, 0%  
 Sequencing vector: plasmid, 100%  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap, version 0.990319  
 Consensus quality: 166139 bases at least Q40  
 Consensus quality: 168833 bases at least Q30  
 Consensus quality: 170280 bases at least Q20  
 Insert size: 185000; agarose-fp  
 Insert size: 175235; sum-of-coverage  
 Quality coverage: 5.37 in Q20 bases; sum-of-coverage  
 Quality coverage: 4.13 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently  
 consists of 21 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 1 1101: contig of 1101 bp in length  
 \* 1102 1201: gap of unknown length  
 \* 1202 2376: contig of 1175 bp in length  
 \* 2377 2476: gap of unknown length  
 \* 2477 5477: contig of 2901 bp in length  
 \* 5478 5477: gap of unknown length  
 \* 5478 7623: contig of 2146 bp in length  
 \* 7624 7723: gap of unknown length

```

* 7724 10114: contig of 2391 bp in length
* 10115 10214: gap of unknown length
* 10215 12550: contig of 2336 bp in length
* 12551 12650: gap of unknown length
* 12651 15385: contig of 2735 bp in length
* 15386 15485: gap of unknown length
* 15486 17583: contig of 2098 bp in length
* 17584 17683: gap of unknown length
* 17684 21196: contig of 3513 bp in length
* 21197 21296: gap of unknown length
* 21297 25327: contig of 4031 bp in length
* 25328 25427: gap of unknown length
* 25428 28981: contig of 3554 bp in length
* 28982 29081: gap of unknown length
* 29082 33331: contig of 4050 bp in length
* 33332 33431: gap of unknown length
* 33432 37028: contig of 3797 bp in length
* 37029 37128: gap of unknown length
* 37129 43068: contig of 5940 bp in length
* 43069 43168: gap of unknown length
* 43169 47557: contig of 4389 bp in length
* 47558 47657: gap of unknown length
* 47658 55576: contig of 7919 bp in length
* 55577 55676: gap of unknown length
* 55677 66582: contig of 10906 bp in length
* 66583 75919: contig of 9237 bp in length
* 75920 94737: gap of unknown length
* 94738 94837: contig of 18718 bp in length
* 94838 128766: gap of unknown length
* 128767 128866: gap of unknown length
* 128867 173519: contig of 44653 bp in length.

FEATURES
source
1..173519
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="UNK"
  /clone="RP24-443G4"
1..1101
  /note="Assembly_name:Contig21"
1202..2376
  /note="Assembly_name:Contig22"
2477..5377
  /note="Assembly_name:Contig24"
5478..7623
  /note="Assembly_name:Contig25"
7724..10114
  /note="Assembly_name:Contig26"
10215..12550
  /note="Assembly_name:Contig27"
12651..15385
  /note="Assembly_name:Contig28"
15486..17583
  /note="Assembly_name:Contig29"
17684..21196
  /note="Assembly_name:Contig30"
21297..25327
  /note="Assembly_name:Contig31"
25428..28981
  /note="Assembly_name:Contig32"
29082..33331
  /note="Assembly_name:Contig33"
33332..37028
  /note="Assembly_name:Contig34"
37129..43068
  /note="Assembly_name:Contig35"
43169..47557
  /note="Assembly_name:Contig36
  clone end:SP6
  vector side:right"
47658..55576
  misc_feature

misc_feature 10114: contig of 2391 bp in length
misc_feature 10214: gap of unknown length
misc_feature 10215 12550: contig of 2336 bp in length
misc_feature 12551 12650: gap of unknown length
misc_feature 12651 15385: contig of 2735 bp in length
misc_feature 15386 15485: gap of unknown length
misc_feature 15486 17583: contig of 2098 bp in length
misc_feature 17584 17683: gap of unknown length
misc_feature 17684 21196: contig of 3513 bp in length
misc_feature 21197 21296: gap of unknown length
misc_feature 21297 25327: contig of 4031 bp in length
misc_feature 25328 25427: gap of unknown length
misc_feature 25428 28981: contig of 3554 bp in length
misc_feature 28982 29081: gap of unknown length
misc_feature 29082 33331: contig of 4050 bp in length
misc_feature 33332 33431: gap of unknown length
misc_feature 33432 37028: contig of 3797 bp in length
misc_feature 37029 37128: gap of unknown length
misc_feature 37129 43068: contig of 5940 bp in length
misc_feature 43069 43168: gap of unknown length
misc_feature 43169 47557: contig of 4389 bp in length
misc_feature 47558 47657: gap of unknown length
misc_feature 47658 55576: contig of 7919 bp in length
misc_feature 55577 55676: gap of unknown length
misc_feature 55677 66582: contig of 10906 bp in length
misc_feature 66583 75919: contig of 9237 bp in length
misc_feature 75920 94737: gap of unknown length
misc_feature 94738 94837: contig of 18718 bp in length
misc_feature 94838 128766: gap of unknown length
misc_feature 128767 128866: gap of unknown length
misc_feature 128867 173519: contig of 44653 bp in length.

BASE COUNT 49702 a 36736 c 34653 g 50398 t 2020 others
ORIGIN
Query Match 74.2%; Score 23; DB 2; length 173519;
Best local similarity 83.9%; Pred. No. 2.1e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ATAGTGAACCTGTGPAATGAAATTTT 31
Db 96369 ATGTGAACCTGTGCAATCTAAATTTT 96369

RESULT 14
AC080138/c
LOCUS
DEFINITION
Mus musculus chromosome 12 clone RP23-361M18 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION
AC080138
VERSION
AC080138.2 GI:11141988
KEYWORDS
HTG; HTGS PHASIS; HTGS DRAFT.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 181191)
Beckstead-Sternberg,S.M., Benjamin,B., Blakeley,R.W.,
Boffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Guan,X.,
Gupta,J., Ho,S.-L., Idol,J., Karlins,E., Lee-Lin,S.-Q., Legaspi,R.,
Lim,M., Maduro,O.L., Maduro,V.B., Masello,C., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
Snyder,B., Stankirpov,S., Thomas,J.W., Thomas,P.J., Tongson,E.E.,
Touchman,J.W., Tsunagawa,C., Vogt,T.L., Walker,M.A., Wehber,K.D.,
Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 181191)
Green,E.D.
Direct Submission
Submitted (26-SEP-2000) NIH Intramural Sequencing Center, 8717
Groveport Circle, Gaithersburg, MD 20877, USA
On Nov 14, 2000 this sequence version replaced gi:10305191.

----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@hghri.nih.gov

----- Project Information
Center project name: sg
Center clone name: 361M18

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 167609 bases at least Q40
Consensus quality: 173188 bases at least Q20
Insert size: 170000; agarose-gel
Insert size: 182000; pulse-field-gel
Insert size: 173291; sum-of-contrigs
Quality coverage: 4.62x in Q20 bases; agarose-gel
Quality coverage: 4.32x in Q20 bases; pulse-field-gel
Quality coverage: 4.38x in Q20 bases; sum-of-contrigs

```

chr	start	end	feature	source		
1	3703	3703	contig of 3703 bp in length	IRRS		
*	3704	3803	gap of unknown length	*		
*	3804	6525	contig of 2822 bp in length	*		
*	6526	6725	gap of unknown length	*		
*	6726	9732	contig of 3007 bp in length	*		
*	9733	9832	gap of unknown length	*		
*	9833	12985	contig of 3153 bp in length	*		
*	12986	13085	gap of unknown length	*		
*	13086	15318	contig of 2233 bp in length	*		
*	15319	15418	gap of unknown length	*		
*	15419	18304	contig of 2886 bp in length	*		
*	18305	18405	gap of unknown length	*		
*	18405	22402	contig of 3598 bp in length	*		
*	22403	22502	gap of unknown length	*		
*	22503	26601	contig of 4099 bp in length	*		
*	26602	26701	gap of unknown length	*		
*	26702	30784	contig of 4083 bp in length	*		
*	30785	30884	gap of unknown length	*		
*	30885	37889	contig of 7005 bp in length	*		
*	37890	43641	gap of unknown length	*		
*	43642	43741	contig of 5652 bp in length	*		
*	43742	43741	gap of unknown length	*		
*	51103	51203	contig of 7362 bp in length	*		
*	51204	55905	gap of unknown length	*		
*	55906	60005	contig of 8702 bp in length	*		
*	60006	65992	gap of unknown length	*		
*	65993	69492	contig of 9387 bp in length	*		
*	69493	81101	gap of unknown length	*		
*	81102	81202	contig of 11609 bp in length	*		
*	81203	92391	gap of unknown length	*		
*	92392	92491	contig of 11190 bp in length	*		
*	92492	108668	gap of unknown length	*		
*	108669	126780	contig of 16177 bp in length	*		
*	126781	126780	gap of unknown length	*		
*	126781	146139	contig of 19559 bp in length	*		
*	146140	181191	gap of unknown length	*		
*	181191	146240	contig of 34952 bp in length	*		
source	Location/Qualifiers					
1	181191	/organism="Mus musculus"				
		/mol_type="genomic DNA"				
		/strain="C57BL6/J"				
		/db_xref="taxon:10090"				
		/chromosome="12"				
		/clone="RP23-351M18"				
		/clone_lib="RP23 mouse BAC library 23"				
	1	3703				
misc_feature	/note="assembly_fragment"					
misc_feature	3804..6625					
misc_feature	/note="assembly_fragment"					
misc_feature	6726..9732					
misc_feature	/note="assembly_fragment"					
misc_feature	9833..12985					
misc_feature	/note="assembly_fragment"					
misc_feature	13086..15318					
misc_feature	/note="assembly_fragment"					
misc_feature	15419..18304					
misc_feature	/note="assembly_fragment"					
misc_feature	18405..22402					
misc_feature	/note="assembly_fragment"					
misc_feature	22503..26601					
misc_feature	/note="assembly_fragment"					

```
misc_feature      /note="assembly_fragment
26702. .30784
clone_end:T7
vector_side:left"
misc_feature      30885. .37889
/note="assembly_fragment"
misc_feature      37990. .43641
/note="assembly_fragment"
43742. .51103
/note="assembly_fragment
clone_end:SP6
vector_side:right"
misc_feature      51204. .59905
/note="assembly_fragment"
60006. .69392
/note="assembly_fragment"
69493. .81101
/note="assembly_fragment"
81202. .92391
/note="assembly_fragment"
92492. .108668
/note="assembly_fragment"
108769. .126680
/note="assembly_fragment"
126781. .146139
/note="assembly_fragment"
146240. .181191
/note="assembly_fragment"
misc_feature      /note="assembly_fragment"
54967 a .34201 c 35187 g 54911 t 1925 others
```

[illegible]

Boutwell, C., Byrne, S., Cantu, C., Castle, A., Cooke, P., Daly, M. J., Depayre, B., Devon, K., Dewar, K., Durette, B., Ettemad, S., Ferreira, P., Forrest, C., Gage, D., Gardina, B., Gensheimer, S., Gierigley, K., Gilmartin, T., Gray, D., Hages, B., Harris, K., Horton, L., Howland, J. C., Hul, L., Jacotot, L., Linton, L., Mackenzie, J., Margule, N., McEwen, P., McGuirk, A., Meldrum, J., Molla, M., Morris, W., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Pavin, B., Peterson, K., Ranganath, S., Riley, R., Roberts, D., Rollins, G., Rossello, R., Roy, A., Shyam, R., Sochoo, S., Stillwell, J., Stone, C., Strickland, C., Spence, K., Tang, L., Vassaliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Ye, W. J., Zemseva, I., Zhao, J. and Zody, M.

**TITLE**  
Direct Submission

**JOURNAL**  
Submitted (03-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 3, 1998 this sequence version replaced gi:2815559.  
The Staden databases, finishing information, and all chromatographic files used in the assembly of this clone are available from our anonymous ftp site.

**COMMENT**

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

**FEATURES**  
source  
Location/Qualifiers  
1..202233  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="hCIT54K19"  
/clone\_1fb="Research Genetics/Cal Tech CITB978SK-B (plates 1-194)"

repeat\_region complement(1447..1509)  
/rpt\_family="AT\_rich"  
complement(1977..2413)  
repeat\_region /rpt\_family="MER28"  
2794..2999  
repeat\_region /rpt\_family="L1"  
3093..3158  
repeat\_region /rpt\_family="AT\_rich"  
3363..3661  
/rpt\_family="AluSg"  
complement(3748..3773)  
repeat\_region /rpt\_family="AT\_rich"  
3828..3899  
repeat\_region /rpt\_family="AT\_rich"  
4086..4148  
repeat\_region /rpt\_family="AT\_rich"  
complement(4318..4397)  
repeat\_region /rpt\_family="LINE2"  
complement(4401..4990)  
repeat\_region /rpt\_family="LINE2"  
complement(5048..5112)  
repeat\_region /rpt\_family="AT\_rich"  
complement(5173..5213)  
repeat\_region /rpt\_family="AT\_rich"  
5572..6874  
repeat\_region /rpt\_family="AluSg"  
complement(7529..8617)  
repeat\_region /rpt\_family="LINE2"  
complement(8992..9065)  
repeat\_region /rpt\_family="MER77"  
complement(9109..9560)  
repeat\_region /rpt\_family="MLT2A"  
complement(9561..9890)  
repeat\_region /rpt\_family="MER77"  
9895..10244  
repeat\_region /rpt\_family="MER33"  
10916..11261  
repeat\_region complement(13745..14445)  
/rpt\_family="L1PAL6"

repeat\_region 14444..14717  
/rpt\_family="L1"  
repeat\_region 14705..15067  
/rpt\_family="L1"  
repeat\_region 15761..15824  
/rpt\_family="MER"  
16004..16827  
repeat\_region /rpt\_family="MER21A"  
complement(17013..17282)  
/rpt\_family="GAAAn"  
complement(17284..17422)  
/rpt\_family="GAn"  
complement(17646..17692)  
/rpt\_family="(CA)n"  
complement(17800..17854)  
/rpt\_family="MLT1F"  
complement(17872..17953)  
/rpt\_family="MLT1F"  
complement(17984..18047)  
/rpt\_family="MSR-INTERNAL"  
complement(18134..18239)  
/rpt\_family="MLT1-INTERNAL"  
complement(18589..18857)  
/rpt\_family="MLT1-INTERNAL"  
complement(18872..19088)  
/rpt\_family="AluSg/X"  
complement(20173..20187)  
/rpt\_family="LINE2"  
complement(20188..20220)  
/rpt\_family="MIR"  
complement(20221..20423)  
/rpt\_family="LINE2"  
complement(20704..20943)  
/rpt\_family="MER69B"  
20946..21236  
repeat\_region /rpt\_family="AluSg"  
complement(21237..21389)  
/rpt\_family="MER69B"  
complement(21390..21679)  
/rpt\_family="AluJo"  
complement(21688..21842)  
/rpt\_family="MER69B"  
complement(21832..21975)  
/rpt\_family="MER69B"  
complement(23805..23903)  
/rpt\_family="L1PAL6"  
complement(23906..24336)  
/rpt\_family="MER57A"  
complement(24337..25040)  
/rpt\_family="L1PAL5"  
25283..26868  
repeat\_region /rpt\_family="L1M4"  
26806..27002  
complement(27028..28892)  
/rpt\_family="L1PAL6"  
28891..29505  
repeat\_region /rpt\_family="L1PAL6"  
29507..30813  
/rpt\_family="L1M3"  
complement(30814..30986)  
/rpt\_family="MSTD"  
30987..31633  
repeat\_region /rpt\_family="L1M3"  
31671..32128  
repeat\_region /rpt\_family="MLT1D"  
32129..33594  
/rpt\_family="L1M3"  
complement(36209..36313)  
/rpt\_family="(GAAA)n"  
complement(36314..36451)  
/rpt\_family="(GAAA)n"  
complement(36501..36787)



=> fil reg; d que l3

FILE 'REGISTRY' ENTERED AT 09:22:13 ON 18 DEC 2003  
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.  
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.  
COPYRIGHT (C) 2003 American Chemical Society (ACS)

Property values tagged with IC are from the ZIC/VINITI data file  
provided by InfoChem.

STRUCTURE FILE UPDATES: 17 DEC 2003 HIGHEST RN 627482-61-5  
DICTIONARY FILE UPDATES: 17 DEC 2003 HIGHEST RN 627482-61-5

TSCA INFORMATION NOW CURRENT THROUGH JULY 14, 2003

Please note that search-term pricing does apply when  
conducting SmartSELECT searches.

Crossover limits have been increased. See HELP CROSSOVER for details.

Experimental and calculated property data are now available. For more  
information enter HELP PROP at an arrow prompt in the file or refer  
to the file summary sheet on the web at:  
<http://www.cas.org/ONLINE/DBSS/registryss.html>

L2 9 SEA FILE=REGISTRY ABB=ON AUUAAGAAGCCCAAGAAAAAUCAGCUGACC|GGUCAG  
CUGAUUUUUCUUGGGCUUCUAU|AUAGUGAAAACUUGUGUAAUUAUGAAAUUUU|AAAAU  
UCAUAAUUACACAAGUUUUCACUAU/SQSN  
L3 4 SEA FILE=REGISTRY ABB=ON L2 AND SQL<101

*Seq 13 2-14  
8-Train  
complements*

=> d rn cn kwic nte lc 1-4

L3 ANSWER 1 OF 4 REGISTRY COPYRIGHT 2003 ACS on STN  
RN 290212-93-0 REGISTRY  
CN GenBank AX028511 (9CI) (CA INDEX NAME)  
SQL 31

SEQ 1 atagtga aaa cttgtgtaat tatgaaattt t  
===== =  
HITS AT: 1-31

\*\*RELATED SEQUENCES AVAILABLE WITH SEQLINK\*\*  
LC STN Files: GENBANK

L3 ANSWER 2 OF 4 REGISTRY COPYRIGHT 2003 ACS on STN  
RN 290212-92-9 REGISTRY  
CN GenBank AX028510 (9CI) (CA INDEX NAME)  
SQL 30

SEQ 1 atatagaagc ccaagaaaaa tcagctgacc  
===== =  
HITS AT: 1-30

\*\*RELATED SEQUENCES AVAILABLE WITH SEQLINK\*\*  
LC STN Files: GENBANK

L3 ANSWER 3 OF 4 REGISTRY COPYRIGHT 2003 ACS on STN  
RN 286027-93-8 REGISTRY  
CN 14: PN: W00042216 SEQID: 14 unclaimed DNA (9CI) (CA INDEX NAME)  
SQL 31

SEQ 1 atagtga aaa cttgtgtaat tatgaaattt t

HITS AT: 1-31

\*\*RELATED SEQUENCES AVAILABLE WITH SEQLINK\*\*

LC STN Files: CA, CAPLUS

L3 ANSWER 4 OF 4 REGISTRY COPYRIGHT 2003 ACS on STN

RN 286027-92-7 REGISTRY

CN 13: PN: W00042216 SEQID: 13 unclaimed DNA (9CI) (CA INDEX NAME)

SQL 30

SEQ 1 atatagaagc ccaagaaaaa tcagctgacc

HITS AT: 1-30

\*\*RELATED SEQUENCES AVAILABLE WITH SEQLINK\*\*

LC STN Files: CA, CAPLUS

=> fil capl; s 13

FILE 'CAPLUS' ENTERED AT 09:22:48 ON 18 DEC 2003

USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.

PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is held by the publishers listed in the PUBLISHER (PB) field (available for records published or updated in Chemical Abstracts after December 26, 1996), unless otherwise indicated in the original publications. The CA Lexicon is the copyrighted intellectual property of the American Chemical Society and is provided to assist you in searching databases on STN. Any dissemination, distribution, copying, or storing of this information, without the prior written consent of CAS, is strictly prohibited.

FILE COVERS 1907 - 18 Dec 2003 VOL 139 ISS 25

FILE LAST UPDATED: 17 Dec 2003 (20031217/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

L4 1 L3

=> d ibib ab hitrn ; fil hom

L4 ANSWER 1 OF 1 CAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2000:493707 CAPLUS

DOCUMENT NUMBER: 133:130737

TITLE: Genetic predisposition to abnormal calcification condition such as osteoporosis

INVENTOR(S): Kusk, Philip

PATENT ASSIGNEE(S): Osteometer Biotech A/S, Den.

SOURCE: PCT Int. Appl., 70 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.

KIND DATE

APPLICATION NO. DATE



-----  
WO 2000042216      A2      20000720      WO 2000-EP319      20000117  
WO 2000042216      A3      20001102  
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,  
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,  
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,  
MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,  
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM,  
AZ, BY, KG, KZ, MD, RU, TJ, TM  
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,  
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG  
EP 1144686      A2      20011017      EP 2000-909076      20000117  
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,  
IE, SI, LT, LV, FI, RO

PRIORITY APPLN. INFO.:

GB 1999-1037      A      19990118  
GB 1999-12585      A      19990528  
WO 2000-EP319      W      20000117

AB      Methods of assessing an individual's predisposition to abnormal  
calcification conditions such as osteoporosis by detg. the genotype of a  
promoter for the bone sialoprotein gene, the matrix gla protein gene, the  
osteopontin gene or the osteoprotegerin gene individually or in any  
combination. Specific allelic variations for each promoter are described.

IT      286027-92-7 286027-93-8

RL: PRP (Properties)

(unclaimed nucleotide sequence; genetic predisposition to abnormal  
calcification condition such as osteoporosis)

FILE 'HOME' ENTERED AT 09:22:55 ON 18 DEC 2003



PI Kusk P;  
 XX  
 DR WPI; 2000-476070/41.  
 XX  
 PT Assessing predisposition to a calcification condition status such as  
 XX osteoporosis and atherosclerosis, comprising determining the genotype  
 PT of the promoter of the bone sialoprotein, matrix gla protein,  
 XX osteopontin or osteoprotegerin gene -  
 XX  
 PS Disclosure; Fig 1; 70pp; English.  
 XX  
 CC The present sequence shows the polymorphic site at position 1869 of the  
 CC human bone sialoprotein (BSP) gene promoter. It was used in the methods  
 CC of the invention, which involve the determination of the promoter  
 CC sequence at polymorphisms in a number of genes controlling bone mineral  
 CC density. These genes include sequences encoding the bone sialoprotein,  
 CC the matrix gla protein (MGP), osteopontin (OPN) and  
 CC osteoprotegerin/osteoclastogenesis inhibitory factor (OPG/OCIF). A at  
 CC this position is associated with a higher peak bone mass. Along with the  
 CC other polymorphic sites, this variable sequence can be used to predict an  
 CC individual's predisposition to osteoporosis and atherosclerosis, thus  
 CC enabling earlier treatment and preventive measures.  
 CC  
 SQ Sequence 31 BP; 12 A; 1 C; 5 G; 13 T; 0 other;  
 Query Match 100.0%; Score 31; DB 21; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 0.1;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ATAGTGAACCTGCTGTAATATGAAATTTT 31  
 1 ATAGTGAACCTGCTGTAATATGAAATTTT 31  
 DB 1 ATAGTGAACCTGCTGTAATATGAAATTTT 31  
 RESULT 2  
 ID AAA61936 standard; DNA; 2421 BP.  
 XX  
 AC AAA61936;  
 XX  
 DT 20-NOV-2000 (first entry)  
 XX  
 DE Human bone sialoprotein (BSP) promoter (-2184-+237).  
 XX  
 KW Bone sialoprotein promoter; BSP; human; tissue-specific promoter;  
 KW targeted expression; therapeutic gene; toxic gene; gene therapy;  
 KW osteolytic tumour; calcification; prostate tumour; osteosarcoma;  
 KW metastatic tumour; lung; breast; colon; brain; multiple myeloma;  
 KW benign prostatic hypertrophy; BPH; arteriosclerosis; osteogenesis;  
 KW osteoblast; bone repair; reporter construct; chromosome 4; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200036919-A1.  
 XX  
 PD 29-JUN-2000.  
 XX  
 PF 22-DEC-1999; 99WO-US30642.  
 XX  
 FR 22-DEC-1998; 98US-0113200.  
 XX  
 PA (UUYI-) UNIV VIRGINIA PATENT FOUND.  
 XX  
 PI Koenneman KS, Chung LMK;  
 XX  
 DR WPI; 2000-442489/38.  
 XX  
 PT Novel therapeutic agent for prevention and treatment of calcified  
 PT tumors and proliferative disorders such as osteosarcoma, multiple  
 PT myeloma and breast cancer, comprises bone sialoprotein promoter -  
 XX  
 PS Claim 5; Fig 8A-B; 79pp; English.  
 XX

CC The invention relates to a novel gene therapy composition comprising a  
 CC mammalian bone sialoprotein (BSP) promoter operably linked to a gene  
 CC sequence encoding a toxic and/or therapeutic protein contained within a  
 CC delivery vector, such as a recombinant adenovirus or a liposome. The BSP  
 CC promoter directs gene expression in tissue and tumour cells with  
 CC calcification potential. It is active in fully-differentiated osteoblasts  
 CC in bone, in trophoblastic cells of the placenta, and in cementum and  
 CC dentin of teeth. It is also active in a variety of human tumours or  
 CC non-malignant growths which exhibit the ability to calcify either in  
 CC the primary or at metastatic sites. This is especially evident for  
 CC prostate and osteosarcoma tumours, but includes any osteotropic  
 CC aggressive metastatic tumour such as, for example, lung, multiple  
 CC myeloma, breast, colon and brain. The BSP promoter may also be expressed  
 CC in non-malignant conditions in which calcification occurs, such as benign  
 CC prostatic hypertrophy (BPH) and arteriosclerosis. The gene therapy  
 CC composition of the invention may be used to treat osteotropic tumours  
 CC (or non-malignant growths with calcification potential) via delivery and  
 CC expression of a gene encoding a protein such as herpes simplex virus  
 CC thymidine kinase (HSVTK), which converts prodrugs such as acyclovir and  
 CC ganciclovir to a cytotoxic phosphorylated form. The composition may also  
 CC be used to promote bone repair via delivery and expression of a gene  
 CC encoding a therapeutic protein such as a growth factor, a cytokine or  
 CC inhibitor thereof, a differentiation factor, a chemokine, an interferon,  
 CC a colony stimulating factor, or an angiogenic factor. The invention  
 CC also discloses a composition in which the gene under the control of  
 CC the BSP promoter is a reporter gene. Such reporter compositions may  
 CC be used to identify compounds capable of modulating osteotropic-specific  
 CC gene expression. The present sequence represents a human BSP promoter  
 CC (-2184-+237) which is claimed for use in the compositions of the  
 CC invention. The human BSP gene is located on chromosome 4.  
 CC  
 SQ Sequence 2421 BP; 891 A; 438 C; 425 G; 667 T; 0 other;  
 Query Match 100.0%; Score 31; DB 21; Length 2421;  
 Best Local Similarity 100.0%; Pred. No. 0.094;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ATAGTGAACCTGCTGTAATATGAAATTTT 31  
 1860 ATAGTGAACCTGCTGTAATATGAAATTTT 1890  
 DB 1860 ATAGTGAACCTGCTGTAATATGAAATTTT 1890  
 RESULT 3  
 ID ABL33972 standard; DNA; 7771 BP.  
 XX  
 AC ABL33972;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SRQ ID NO: 1945.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cyostatic; nocrotropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-BP07537.  
 XX  
 FR 30-JUN-2000; 2000DE-1032529.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX

XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX  
XX Claim 1; SEQ ID NO 1945; 32bp + Sequence Listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
XX Sequence 7771 BP; 2788 A; 45 C; 1303 G; 3635 T; 0 other;  
SQ  
XX  
XX Query Match 89.7%; Score 27.8; DB 24; Length 7771;  
XX Best Local Similarity 93.5%; Pred. No. 1.1;  
XX Matches 29; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 ATAGTGAACCTGTGTATATGAAATTT 31  
DB 4650 ATAGTGAACCTGTGTATATGAAATTT 4680  
XX  
XX RESULT 4  
XX ABL33973/c.  
XX ID ABL33973 standard; DNA; 7771 BP.  
XX  
XX ABL33973;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Human immune system associated gene SEQ ID NO: 1946.  
XX  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; anti-amyloid; cytosine methylation; nocturnal;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200200928-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001MO-EP07537.  
XX  
XX 30-JUN-2000; 2000DE-103529.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIDEMIOLOGY AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX  
XX Claim 1; SEQ ID NO 1946; 32bp + Sequence Listing; German.  
XX

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
XX Sequence 7771 BP; 2143 A; 45 C; 1537 G; 4046 T; 0 other;  
SQ  
XX  
XX Query Match 74.2%; Score 23; DB 24; Length 7771;  
XX Best Local Similarity 83.9%; Pred. No. 50;  
XX Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
OY 1 ATAGTGAACCTGTGTATATGAAATTT 31  
DB 3122 ATAGTGAACCTGTGTATATGAAATTT 3092  
XX  
XX RESULT 5  
XX ABX26253  
XX ID ABX26253 standard; cDNA; 134 BP.  
XX  
XX ABX26253;  
XX  
XX 11-FEB-2003 (first entry)  
XX  
XX Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #8310.  
XX  
XX Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;  
KW cellular fucosylation; glycosyltransferase; fucosylation; transplant rejection;  
KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;  
KW complex carbohydrate; gene replacement therapy; immunosuppressive;  
KW antiinflammatory; antiarthritic; antibacterial; cerebroprotective;  
KW antiasthmatic; vasotrophic.  
XX  
XX Homo sapiens.  
OS  
XX US2002110548-A1.  
XX  
XX 15-AUG-2002.  
XX  
XX 11-JUN-2001; 2001US-0878574.  
XX  
XX 22-NOV-1996; 96US-0753233.  
XX  
XX 03-DEC-1997; 97US-0984246.  
XX  
XX 09-SEP-1998; 98US-0149674.  
XX  
XX 14-JUN-1999; 99US-0333177.  
XX  
XX (GENY) GENETICS INST INC.  
XX  
XX Sullivan F, Kriz R, Kumar R;  
XX  
XX WPI; 2003-066673/06.  
XX  
XX New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)  
PT peptide, for manufacturing complex carbohydrates, or as targets for  
PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant  
PT rejection  
XX  
XX Disclosure; SEQ ID NO 8312; 6pp; English.  
XX  
XX The invention relates to a composition comprising a human GDP-mannose  
CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying  
CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation  
CC in a mammalian subject and for treating or ameliorating diseases affected  
CC by the level of cellular fucosylation or diseases affected by the  
CC fucosylation of glycoconjugates. These diseases include arthritis,  
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or  
CC infection. The GM4,6D peptide or a polynucleotide encoding it is also  
CC useful for manufacturing complex carbohydrates and as targets for  
CC screening small molecule antagonists of the activity of the enzyme. The

CC polynucleotide is useful in developing an assay for defects in the  
CC enzyme, as well as in gene replacement therapy. Sequences  
CC ABX17942-ABX17944 and ABX17947-ABX3716 represent DNA molecules encoding  
CC human GM4, 6D peptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from USPTO  
CC at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).

XX SQ Sequence 134 BP; 40 A; 13 C; 25 G; 56 T; 0 other;

Query Match 72.9%; Score 22.6; DB 25; Length 134;  
Best Local Similarity 85.2%; Pred. No. 74;  
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 TAGTGAACCTGCTGTAATTAATTAATTT 30  
DB 39 TAGTGAACCTGCTGTAATTAATTAATTT 67

RESULT 6

AXX61630/c  
ID AAX61630 standard; DNA; 615 BP.

XX AC AAX61630;

XX DT 19-JUL-1999 (first entry)

XX DE B. burgdorferi antigenic protein coding sequence, f739.nt.

XX KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

XX KM Borrelia burgdorferi.

XX OS WO9859071-AL.

XX PD 30-DEC-1998.

XX PF 18-JUN-1998; 98WO-US12718.

XX PR 03-SEP-1997; 97US-0057483.

XX PR 20-JUN-1997; 97US-0050359.

XX PR 22-JUL-1997; 97US-0053344.

XX PR 22-JUL-1997; 97US-0053377.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (MEDI-) MEDIMUNE INC.

XX P1 Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX DR MPI, 1999-189980/16.

XX DR P-PSDB; AAY19933.

XX PT New isolated Borrelia burgdorferi nucleic acids - used to develop

XX PT products for the diagnosis, prevention and treatment of diseases

XX PT caused by Borrelia, particularly Lyme disease

XX PS Claim 1; Page 124; 275pp; English.

XX CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the

XX CC invention, which is suitable for use in a vaccine. The Bb polypeptides

XX CC can be used in vaccines for eliciting protective antibodies to members of

XX CC the Borrelia genus, particularly for the use against Lyme disease in

XX CC humans and animals. They can be used for preventing or attenuating an

XX CC infection caused by a member of the Borrelia genus. The products can also

XX CC be used for detection of members of the Borrelia genus.

DB 576 ATATATTAACCTGCTGTAATTAATTAAT 549

RESULT 7

AXX61629/c

ID AAX61629 standard; DNA; 654 BP.

XX AC AAX61629;

XX DT 19-JUL-1999 (first entry)

XX DE B. burgdorferi antigenic protein coding sequence, f739.nt.

XX KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

XX KM Borrelia burgdorferi.

XX OS WO9859071-AL.

XX PD 30-DEC-1998.

XX PF 18-JUN-1998; 98WO-US12718.

XX PR 03-SEP-1997; 97US-0057483.

XX PR 20-JUN-1997; 97US-0050359.

XX PR 22-JUL-1997; 97US-0053344.

XX PR 22-JUL-1997; 97US-0053377.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (MEDI-) MEDIMUNE INC.

XX P1 Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX DR MPI, 1999-189980/16.

XX DR P-PSDB; AAY19932.

XX PT New isolated Borrelia burgdorferi nucleic acids - used to develop

XX PT products for the diagnosis, prevention and treatment of diseases

XX PT caused by Borrelia, particularly Lyme disease

XX PS Claim 1; Page 124; 275pp; English.

XX CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the

XX CC invention, which is suitable for use in a vaccine. The Bb polypeptides

XX CC can be used in vaccines for eliciting protective antibodies to members of

XX CC the Borrelia genus, particularly for the use against Lyme disease in

XX CC humans and animals. They can be used for preventing or attenuating an

XX CC infection caused by a member of the Borrelia genus. The products can also

XX CC be used for detection of members of the Borrelia genus.

XX SQ Sequence 654 BP; 227 A; 61 C; 104 G; 262 T; 0 other;

Query Match 69.7%; Score 21.6; DB 20; Length 654;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ATATATTAACCTGCTGTAATTAATTAAT 28  
DB 615 ATATATTAACCTGCTGTAATTAATTAAT 588

RESULT 8

AXX90942

ID AAX90942 standard; DNA; 7737 BP.

XX AC AAX90942;

XX DT 05-NOV-2001 (first entry)

XX DE Human digestive system antigen genomic sequence SEQ ID NO: 4518.

XX DE Human digestive system antigen; gene therapy; cancer; appendicitis;

XX KM Human; digestive system antigen; gene therapy; cancer; appendicitis;

KM	ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX	digestive system disorder; Meckel's diverticulum; ds.
XX	Homo sapiens.
OS	MO200155314-A2.
XX	
PM	02-AUG-2001.
XX	
PD	17-JAN-2001; 2001WO-US01334.
PF	
XX	
XX	31-JAN-2000; 2000US-0179065.
XX	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198113.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216637.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225265.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226688.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227909.
PR	30-AUG-2000; 2000US-0228924.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.
PR	08-SEP-2000; 2000US-0231243.
PR	08-SEP-2000; 2000US-0231244.
PR	08-SEP-2000; 2000US-0231413.
PR	08-SEP-2000; 2000US-0231414.
PR	08-SEP-2000; 2000US-0232080.
PR	12-SEP-2000; 2000US-0232081.
PR	12-SEP-2000; 2000US-0231968.
PR	14-SEP-2000; 2000US-0232357.
PR	14-SEP-2000; 2000US-0232359.
PR	14-SEP-2000; 2000US-0232399.
PR	14-SEP-2000; 2000US-0232400.
PR	14-SEP-2000; 2000US-0232401.
PR	14-SEP-2000; 2000US-0233063.
PR	14-SEP-2000; 2000US-0233064.
PR	14-SEP-2000; 2000US-0233065.
PR	21-SEP-2000; 2000US-0234223.
PR	21-SEP-2000; 2000US-0234224.
PR	25-SEP-2000; 2000US-0234297.
PR	25-SEP-2000; 2000US-0234298.
PR	26-SEP-2000; 2000US-0234584.
PR	27-SEP-2000; 2000US-0235834.
PR	27-SEP-2000; 2000US-0235835.
PR	29-SEP-2000; 2000US-0236327.
PR	29-SEP-2000; 2000US-0236367.
PR	29-SEP-2000; 2000US-0236368.
PR	29-SEP-2000; 2000US-0236370.
PR	29-SEP-2000; 2000US-0236370.
PR	02-OCT-2000; 2000US-0236802.
PR	02-OCT-2000; 2000US-0237037.
PR	02-OCT-2000; 2000US-0237038.
PR	02-OCT-2000; 2000US-0237039.
PR	02-OCT-2000; 2000US-0237040.
PR	13-OCT-2000; 2000US-0239935.
PR	13-OCT-2000; 2000US-0239937.
PR	20-OCT-2000; 2000US-0240960.
PR	20-OCT-2000; 2000US-0241826.
PR	20-OCT-2000; 2000US-0241827.
PR	20-OCT-2000; 2000US-0241785.
PR	20-OCT-2000; 2000US-0241786.
PR	20-OCT-2000; 2000US-0241787.
PR	20-OCT-2000; 2000US-0241808.
PR	20-OCT-2000; 2000US-0241809.
PR	20-OCT-2000; 2000US-0241826.
PR	01-NOV-2000; 2000US-0244617.
PR	08-NOV-2000; 2000US-0246474.
PR	08-NOV-2000; 2000US-0246475.
PR	08-NOV-2000; 2000US-0246476.
PR	08-NOV-2000; 2000US-0246477.
PR	08-NOV-2000; 2000US-0246478.
PR	08-NOV-2000; 2000US-0246523.
PR	

PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-502630/55.  
XX  
XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
XX Disclosure; SEQ ID NO 4518; 986bp; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
CC encoding a digestive system antigen of the invention.  
XX  
XX Sequence 7737 BP; 2397 A; 1325 C; 1490 G; 2525 T; 0 other;  
SQ  
Query Match 69.7%; Score 21.6; DB 22; Length 7737;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
CY 4 GTGAAACTGTGTATATATGAAATTTT 31  
DB 4085 GTGAACATGTGTATATATGAAATTTT 4112  
RESULT 9  
AAS31977  
ID AAS31977 standard; DNA: 7737 BP.  
XX  
XX AAS31977;  
XX  
XX 04-DEC-2001 (first entry)  
XX  
XX Human liver associated genomic DNA #151.  
DE  
XX Liver associated protein; human; mouse; rabbit; goat; horse; cat; dog;  
KW chicken; sheep; immunosuppressive; antidiarrhetic; vasotrophic;  
KW antineoplastic; antiproliferative; cytostatic; cardiac; neuroprotective;  
KW cerebroprotective; neurotropic; antibacterial; virucide; fungicide; cancer;  
KW ophthalmological; venereal; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; tissue regeneration;  
KW anti-fertility.  
XX  
XX Homo sapiens.  
OS  
XX WO200155355-A1.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01351.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 26-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225265.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227609.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0228927.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0233397.  
PR 14-SEP-2000; 2000US-0233398.  
PR 14-SEP-2000; 2000US-0233399.  
PR 14-SEP-2000; 2000US-0233400.  
PR 14-SEP-2000; 2000US-0233401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0251989.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251859.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-457728/49.  
XX  
XX Isolated nucleic acid molecule encoding a human liver related protein  
XX is used in preventing, treating or ameliorating disorders of the liver  
XX particularly cancer of the liver -  
XX  
XX Claim 1; SEQ ID No 453; 526bp; English.  
XX  
XX Sequences AAS31827-AAS32182 represent genomic DNA molecules, which encode  
XX the liver associated polypeptides of the invention. Liver associated  
XX polypeptides and their associated polynucleotides are useful in the  
XX diagnosis, treatment and prevention of various types of disorders in e.g.  
XX humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A  
XX pathological condition can be determined by detecting the presence or  
XX absence of a mutation in a liver associated polynucleotide. The treatable

CC disorders include autoimmune diseases such as rheumatoid arthritis,  
CC hyperproliferative disorders such as neoplasms of the breast or liver,  
CC cardiovascular disorders such as cardiac arrest, cerebrovascular  
CC disorders such as cerebral ischaemia, nervous system disorders such as  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,  
CC ocular disorders such as corneal infection, endocrine disorders such as  
CC premature labour and infertility, gastrointestinal disorders such as  
CC Crohn's disease, renal disorders such as glomerulonephritis and  
CC respiratory disorders such as asthma and pleurisy. The polypeptides can  
CC also be used to aid wound healing, to prevent skin aging due to sunburn,  
CC to maintain organs before transplantation, to regenerate tissues and in  
CC chemotaxis.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

Query Match 69.7%; Score 21.6; DB 22; Length 7737;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 GTGAACCTTGTAATATGAATTTT 31  
Db 4085 GTGAACCTTGTAATATGAATTTT 4112

RESULT 10  
ABK43114  
ID ABK43114 standard; DNA; 7737 BP.  
XX  
XX AC ABK43114;  
XX  
XX

DT 21-MAY-2002 (first entry)

DE Genomic sequence #1013 encoding human connective tissue polypeptide.

XX  
XX KW Human; connective tissue related disorder; cancer; gene therapy;  
XX cytostatic; gene; ds.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WC00155343-A1.  
XX  
XX PD 02-AUG-2001.  
XX  
XX PF 17-JAN-2001; 2001WO-US01322.  
XX  
XX

XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0186874.  
XX 17-MAR-2000; 2000US-0198123.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205115.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 14-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220863.  
XX 26-JUL-2000; 2000US-0220864.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.



```

PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250360.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251890.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM,
XX WPI, 2001-565190/63.
XX
XX Nucleic acid encoding novel connective tissue associated polypeptides,
XX used in diagnosing, preventing, treating or ameliorating a disorder
XX such as cancer or rheumatoid arthritis -
XX
XX Disclosure, SEQ ID No 2001; 673bp; English.
XX
XX CC The present invention relates to the isolation of novel human connective
XX CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
XX CC (cDNA and genomic) sequences encoding them. The sequences of the
XX CC invention are useful in the diagnosis, treatment, prevention and/or
XX CC prognosis of diseases associated with connective tissue(s), including
XX CC cancer. The polynucleotide sequences of the invention are also useful
XX CC in gene therapy. ABK43102-ABK43116 represent genomic sequences encoding
XX CC the novel human connective tissue related polypeptides.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 7737 BP; 2397 A; 1325 C; 1490 G; 2525 T; 0 other;

Query Match 69.7%; Score 21.6; DB 23; Length 7737;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

CY 4 GTGAAACTGTGTATTAATGCAATTTT 31
Db 4085 GTGAAACTGTGTATTAATTAATTTT 4112

```

PR	01-NOV-2000;	2000US-244617P.
PR	17-NOV-2000;	2000US-249299P.
PR	08-DEC-2000;	2000US-251856P.
PR	08-DEC-2000;	2000US-251868P.
PR	08-DEC-2000;	2000US-251869P.
XX	(ROSE/)	ROSEN C A.
PA	(RUBEN/)	RUBEN S M.
PA	(BARA/)	BARASH S C.
XX		
XX	Rosen CA,	Ruben SM, Barash SC;
XX		
DR	WPI;	2002-381944/41.
PT	New nucleic acid encoding human liver antigens, useful for diagnosis,	
PT	treatment and prevention of e.g. hepatitis and hepatic cancer, also	
PT	related polypeptides and antibodies	
XX		
PS	Disclosure; SEQ ID NO 453; 181bp; English.	
XX		
CC	The invention relates to 145 novel human liver antigens (ABP40831-	
CC	ABP40975) and to cDNAs encoding them (ABN90036-ABN90180), and also	
CC	encompasses polypeptides 90% identical and polynucleotides 95% identical	
CC	to the sequences of the invention. The invention additionally relates to	
CC	recombinant vectors and host cells comprising human liver antigen	
CC	polynucleotides, antibodies against human liver antigens, and the use of	
CC	liver antigen polynucleotides and polypeptides in diagnosing, treating,	
CC	prophesing or preventing various disorders of the liver. Such conditions	
CC	include viral infections (e.g., cytomegalovirus, Epstein-Barr virus,	
CC	hepatitis A virus, hepatitis B virus and hepatitis C virus), parasitic	
CC	infections (e.g., Clonorchis sinensis, Echinococcus granulosus and	
CC	Entamoeba histolytica), and also bacterial and fungal infections. Other	
CC	disorders that may be treated include inflammatory conditions (e.g.,	
CC	cirrhosis and granulomatous hepatitis), damage caused by drugs or toxins,	
CC	autoimmune diseases (e.g. Wilson's disease, primary biliary cirrhosis),	
CC	neoplastic disorders (e.g., adenomas, haemangiomas and hepatocellular	
CC	carcinoma), portal hypertension, or gastrointestinal disorders (e.g.,	
CC	peptic ulcers, gastritis and peritoneal diseases). Liver antigen	
CC	polypeptides and polynucleotides may also be used in screening for	
CC	compounds which modulate liver antigen expression or activity. The	
CC	polynucleotides may further be used for gene therapy, chromosome	
CC	mapping, in the identification of individuals and in forensic analysis,	
CC	and the polypeptides may be used as molecular weight markers or to	
CC	prepare antibodies useful in disease diagnosis, drug targeting and	
CC	phenotyping. Sequences ABN90182-ABN90537 represent human liver antigen	
CC	genomic sequences.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from the	
CC	USPTO at segdata.uspto.gov/sequence/	
XX		
SQ	Sequence 7737 BP; 2397 A; 1325 C; 1490 G; 2525 T; 0 other;	
	Query Match	69.7%; Score 21.6; DB 24; Length 7737;
	Best Local Similarity	85.7%; Pred. No. 1.5e+02;
	Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0.	
OY	4 GTGAACCTGTCTAATTCATGCAATTTT 31	
DB		
	4085 GTGAACCTGTCTAATTCATGCAATTTT 4112	
RESUME 12		
AA546476	AA546476 standard; DNA; 13377 BP.	
XX	AA546476;	
XX		
DT	18-DEC-2001 (first entry)	
XX		
DE	Tumour suppressor gene derived chemically modified sequence #198.	
XX		
KW	Human; tumour suppressor gene; oncogene; antitumour; cytostatic;	
KW	cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;	

KW cytosine methylation; ds.  
 OS Homo sapiens.  
 PN WO200168912-A2.  
 XX 20-SEP-2001.  
 XX 15-MAR-2001; 2001WO-BP02955.  
 XX 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032522.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX (EPiG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K.  
 XX WPI; 2001-602752/68.  
 XX Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for  
 PT analysing diseases associated with cytosine methylation state e.g.  
 PT cancer  
 XX  
 XX Claim 1; SEQ ID NO 198; 27bp; English.  
 CC The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes, having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a  
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with CpG dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC and/or therapy of existing diseases or the predisposition to specific  
 CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and  
 CC oncogenes. Sequences with even numbered Seq ID numbers are the  
 CC complementary sequence of the corresponding odd numbered sequence (e.g.  
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
 CC is missing).  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 SQ Sequence 13377 BP; 3714 A; 146 C; 2702 G; 6815 T; 0 other;  
 Query Match 69.7%; Score 21.6; DB 22; Length 13377;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 4 GTGAAGAACTTGTGTAATTATGAAATTTT 31  
 DB 1827 GTGAAGAACTTGTGTAATTATGAAATTTT 1854  
 RESULT 13  
 ABL33463  
 ID ABL33463 standard; DNA; 13377 BP.  
 XX  
 AC ABL33463;  
 XX

DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 1436.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cytosatic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antineumatic; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; de.  
 XX  
 OS Homo sapiens.  
 PN WO200200928-A2.  
 XX 03-JAN-2002.  
 XX 02-JUL-2001; 2001WO-BP07537.  
 XX 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX (EPiG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K.  
 XX WPI; 2002-130909/17.  
 DR Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation  
 XX  
 XX Claim 1; SEQ ID NO 1436; 32bp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 SQ Sequence 13377 BP; 3714 A; 146 C; 2702 G; 6815 T; 0 other;  
 Query Match 69.7%; Score 21.6; DB 24; Length 13377;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 4 GTGAAGAACTTGTGTAATTATGAAATTTT 31  
 DB 1827 GTGAAGAACTTGTGTAATTATGAAATTTT 1854  
 RESULT 14  
 AAX20250/C  
 ID AAX20250 standard; DNA; 111309 BP.  
 XX  
 AC AAX20250;  
 XX 04-MAY-1999 (first entry)  
 DT  
 XX  
 DE Borrelia burgdorferi polymucleotide sequence #3.  
 XX  
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.  
 XX  
 OS Borrelia burgdorferi.  
 XX  
 PN W09858943-A1.

```

XX 30-DEC-1998.
PD
XX
PF 18-JUN-1998; 98WO-US12764.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI (MED1-) MEDIMMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI White OR;
XX
DR WPI; 1999-081217/07.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
XX
XX Claim 1; Page 738-800; 1128pp; English.
XX
CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
XX
SQ Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T; 10 other;
XX
Query Match 69.7%; Score 21.6; DB 20; Length 111309;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
QY 1 ATAGTGAAGAACTGTGATATGATGAAT 28
DB 102865 ATAAATGAAGAACTGTGATATGATGAAT 102838
XX
RESULT 15
AAX20248/c
ID AAX20248 standard; DNA; 910715 BP.
XX
AC AAX20248;
XX
DT 04-MAY-1999 (first entry)
XX
DE Borrelia burgdorferi polynucleotide sequence #1.
XX
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
XX
OS Borrelia burgdorferi.
XX
PN MO9858943-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12764.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

```

```

PA (MED1-) MEDIMMUNE INC.
XX
XX Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI White OR;
XX
DR WPI; 1999-081217/07.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
XX
XX Claim 1; Page 157-671; 1128pp; English.
XX
CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
XX
SQ Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;
XX
Query Match 69.7%; Score 21.6; DB 20; Length 910715;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
QY 1 ATAGTGAAGAACTGTGATATGATGAAT 28
DB 692251 ATAAATGAAGAACTGTGATATGATGAAT 692224
XX

```

Search completed: December 18, 2003, 10:30:05  
Job time : 212.279 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2003, 09:02:17 ; Search time 45.2295 seconds  
(without alignments)  
302.521 Million cell updates/sec

Title: US-09-889-491-14

Perfect score: 1 atagctgaactctgtcaattatgaatttt 31

Scoring table: IDENTITY NUC  
Gapop 10-0, Gapext 1-0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:  
5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq:  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	62.6	1001	US-09-641-638-465	Sequence 465, App
2	19.4	62.6	1001	US-09-641-638-466	Sequence 466, App
3	19.4	62.6	1002	US-09-641-638-595	Sequence 595, App
4	19.4	62.6	1558	US-09-286-691-1	Sequence 1, Appli
5	19.4	62.6	1558	US-09-687-147-1	Sequence 117, App
6	19.4	62.6	3001	US-09-539-333D-117	Sequence 216, App
7	19.4	62.6	3001	US-09-539-333D-216	Sequence 1, Appli
8	19.2	61.9	162450	US-09-345-882-1	Sequence 19, Appli
9	19.2	61.3	14855	US-08-687-080-59	Sequence 58, Appli
10	18.8	60.6	594	PCT-US96-07709-18	Sequence 20, Appli
11	18.8	60.6	594	PCT-US96-07709-20	Sequence 27, Appli
12	18.8	60.6	744	PCT-US96-07709-27	Sequence 28, Appli
13	18.8	60.6	744	PCT-US96-07709-28	Sequence 26, Appli
14	18.8	60.6	905	PCT-US96-07709-24	Sequence 29, App
15	18.8	60.6	905	PCT-US96-07709-26	Sequence 29, App
16	18.8	60.6	4503	US-09-620-312D-293	Sequence 29, App
17	18.8	60.6	4617	US-09-620-312D-294	Sequence 29, App
18	18.8	60.6	11789	US-09-387-286-57	Sequence 87, Appli
19	18.8	60.6	12665	US-08-961-527-134	Sequence 134, App
20	18.8	60.6	580073	US-08-545-528D-1	Sequence 1, Appli
21	18.8	60.6	1664976	US-08-916-418-1	Sequence 1, Appli
22	18.6	60.0	116592	US-09-818-512-3	Sequence 3, Appli
23	18.4	59.4	500	US-08-967-101-86	Sequence 86, Appli
24	18.4	59.4	500	US-08-592-541-86	Sequence 86, Appli
25	18.4	59.4	500	US-09-124-698-86	Sequence 86, Appli
26	18.4	59.4	500	US-09-127-480-86	Sequence 86, Appli
27	18.4	59.4	500	US-08-496-841C-86	Sequence 86, Appli

#### ALIGNMENTS

28	18.4	59.4	500	4	US-09-124-523-86	Sequence 86, Appli
29	18.4	59.4	500	4	US-09-636-796A-86	Sequence 86, Appli
30	18.4	59.4	500	4	US-08-431-048F-86	Sequence 86, Appli
31	18.4	59.4	1581	4	US-09-601-158-80	Sequence 80, Appli
32	18.4	59.4	2451	4	US-09-328-352-2551	Sequence 2551, Ap
33	18.4	59.4	1230025	4	US-09-198-452A-1	Sequence 1, Appli
34	18.2	58.7	717	4	US-09-328-352-1778	Sequence 1778, Ap
35	18.2	58.7	1419	1	US-07-960-981-4	Sequence 4, Appli
36	18.2	58.7	1419	5	PCT-US93-09634-4	Sequence 4, Appli
37	18.2	58.7	1456	4	US-09-064-411A-23	Sequence 23, Appli
38	18.2	58.7	2539	3	US-08-749-522-3	Sequence 3, Appli
39	18.2	58.7	3794	4	US-09-193-434-1	Sequence 1, Appli
40	18.2	58.7	4177	3	US-09-023-082A-23	Sequence 23, Appli
41	18.2	58.7	4379	3	US-08-593-214A-17	Sequence 17, Appli
42	18.2	58.7	4379	3	US-09-149-976-17	Sequence 17, Appli
43	18.2	58.7	7210	2	US-08-1257-963B-10	Sequence 10, Appli
44	18.2	58.7	7210	4	US-08-367-841A-10	Sequence 10, Appli
45	18.2	58.7	7210	5	PCT-US95-07201-10	Sequence 10, Appli

#### RESULT 1

US-09-641-638-465/c  
Sequence 465, Application US/09641638  
Patent No. 6432648  
GENERAL INFORMATION:  
APPLICANT: Blumentfeld, Marta  
APPLICANT: Bouguetelret, Lydie  
APPLICANT: Chumakov, Ilya  
APPLICANT: Cohen, Amich  
TITLE OF INVENTION: BIOMETRIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
FILE REFERENCE: GENSET.051CEP  
CURRENT FILING DATE: US/09/641,638  
PRIOR APPLICATION NUMBER: US/09/502,330  
PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: US 60/133,200  
PRIOR FILING DATE: 1999-05-07  
PRIOR APPLICATION NUMBER: US 09/275,267  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: US 60/119,917  
PRIOR FILING DATE: 1999-02-12  
NUMBER OF SEQ ID NOS: 1304  
SOFTWARE: Patent.pm  
SEQ ID NO 465  
LENGTH: 1001  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 501  
OTHER INFORMATION: 12-44-50 : polymorphic base T or C  
NAME/KEY: misc binding  
LOCATION: 502..521  
OTHER INFORMATION: 12-44-50.misl, potential complement  
NAME/KEY: misc binding  
LOCATION: 482..500  
OTHER INFORMATION: 12-44-50.mis2  
NAME/KEY: primer bind  
LOCATION: 530..550  
OTHER INFORMATION: upstream amplification primer, complement  
NAME/KEY: primer bind  
LOCATION: 147..166  
OTHER INFORMATION: downstream amplification primer  
NAME/KEY: misc binding  
LOCATION: 489..513  
OTHER INFORMATION: 12-44-50 potential probe  
US-09-641-638-465  
Query Match 62.6%; Score 19.4; DB 4; Length 1001;

Best Local Similarity 79.3%; Pred. No. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATAGTGAACCTGTGTATATGAAATT 29  
Db 863 ATAGTGAACCTGTGTATATGAAATT 835

## RESULT 2

US-09-641-638-466/c

Sequence 466, Application US/09641638  
Patent No. 6432648

## GENERAL INFORMATION:

APPLICANT: Blumenfeld, Marta  
APPLICANT: Bougueleret, Lydie

APPLICANT: Chumakov, Ilya

TITLE OF INVENTION: BIALLIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM

FILE REFERENCE: GENSET.051CPI

CURRENT APPLICATION NUMBER: US/09/641,638

PRIOR FILING DATE: 2000-08-16

PRIOR APPLICATION NUMBER: US 09/502,330

PRIOR FILING DATE: 2000-02-11

PRIOR APPLICATION NUMBER: US 60/133,200

PRIOR FILING DATE: 1999-05-07

PRIOR APPLICATION NUMBER: US 09/275,267

PRIOR FILING DATE: 1999-03-23

PRIOR APPLICATION NUMBER: US 60/119,917

PRIOR FILING DATE: 1999-02-12

NUMBER OF SEQ ID NOS: 1304

SOFTWARE: Patent.pm

SEQ ID NO 466

LENGTH: 1001

TYPE: DNA

ORGANISM: Homo Sapiens

FEATURE:

NAME/KEY: allele

LOCATION: 501

OTHER INFORMATION: 12-44-67 : polymorphic base T or C

NAME/KEY: misc binding

LOCATION: 481..500

OTHER INFORMATION: 12-44-67, misl, potential

NAME/KEY: misc binding

LOCATION: 502..521

OTHER INFORMATION: 12-44-67, mis2, potential complement

NAME/KEY: primer bind

LOCATION: 547..567

OTHER INFORMATION: upstream amplification primer, complement

NAME/KEY: primer bind

LOCATION: 164..183

OTHER INFORMATION: downstream amplification primer

NAME/KEY: misc binding

LOCATION: 489..513

OTHER INFORMATION: 12-44-67 potential probe

US-09-641-638-466

Query Match 62.6%; Score 19.4; DB 4; Length 1001;  
Best Local Similarity 79.3%; Pred. No. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATAGTGAACCTGTGTATATGAAATT 29  
Db 880 ATAGTGAACCTGTGTATATGAAATT 852

## RESULT 3

US-09-641-638-595/c

Sequence 595, Application US/09641638  
Patent No. 6432648

## GENERAL INFORMATION:

APPLICANT: Blumenfeld, Marta

APPLICANT: Bougueleret, Lydie

APPLICANT: Chumakov, Ilya  
APPLICANT: Cohen, Annick  
TITLE OF INVENTION: BIALLIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM

FILE REFERENCE: GENSET.051CPI

CURRENT APPLICATION NUMBER: US/09/641,638

CURRENT FILING DATE: 2000-08-16

PRIOR APPLICATION NUMBER: US 09/502,330

PRIOR FILING DATE: 2000-02-11

PRIOR APPLICATION NUMBER: US 60/133,200

PRIOR FILING DATE: 1999-05-07

PRIOR APPLICATION NUMBER: US 09/275,267

PRIOR FILING DATE: 1999-03-23

PRIOR APPLICATION NUMBER: US 60/119,917

PRIOR FILING DATE: 1999-02-12

NUMBER OF SEQ ID NOS: 1304

SOFTWARE: Patent.pm

SEQ ID NO 595

LENGTH: 1002

TYPE: DNA

ORGANISM: Homo Sapiens

FEATURE:

NAME/KEY: allele

LOCATION: 458

OTHER INFORMATION: 12-44-181 : deletion C

NAME/KEY: misc binding

LOCATION: 438..457

OTHER INFORMATION: 12-44-181, misl, potential

NAME/KEY: primer bind

LOCATION: 618..638

OTHER INFORMATION: upstream amplification primer, complement

NAME/KEY: primer bind

LOCATION: 235..254

OTHER INFORMATION: downstream amplification primer

US-09-641-638-595

Query Match 62.6%; Score 19.4; DB 4; Length 1002;  
Best Local Similarity 79.3%; Pred. No. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATAGTGAACCTGTGTATATGAAATT 29  
Db 951 ATAGTGAACCTGTGTATATGAAATT 923

## RESULT 4

US-09-286-691-1

Sequence 1, Application US/09286691  
Patent No. 6190189

## GENERAL INFORMATION:

APPLICANT: Li, Xin-Liang

APPLICANT: Jungdahl, Lars G.

APPLICANT: Chen, Huizhong

TITLE OF INVENTION: Cellulases and Coding Sequences

FILE REFERENCE: 42-96

CURRENT APPLICATION NUMBER: US/09/286,691

CURRENT FILING DATE: 1999-04-05

EARLIER APPLICATION NUMBER: US 60/027,883

EARLIER FILING DATE: 1996-10-04

EARLIER APPLICATION NUMBER: PCT US97/18008

EARLIER FILING DATE: 1997-10-03

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1

LENGTH: 1558

TYPE: DNA

ORGANISM: Orpiniomyces sp. PC-2

FEATURE:

NAME/KEY: CDS

LOCATION: (105)..(1481)

US-09-286-691-1

Query Match 62.6%; Score 19.4; DB 3; Length 1558;

Best Local Similarity 79.3%; Pred. No. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TAGGGAACCTTGCTGATATGAAATTT 30  
DB 491 TAGGATACCTCTTGGAAATGAAATTT 519

## RESULT 5

US-09-687-147-1  
Sequence 1, Application US/09687147

Patent No. 6268198

GENERAL INFORMATION:

APPLICANT: Li, Xian-Liang

APPLICANT: Ljungdahl, Lars G.

APPLICANT: Chen, Huizhong

TITLE OF INVENTION: Cellulases and Coding Sequences

FILE REFERENCE: 42-96a

CURRENT APPLICATION NUMBER: US/09/687,147

CURRENT FILING DATE: 2000-10-12

PRIOR APPLICATION NUMBER: US 60/027,883

PRIOR FILING DATE: 1996-10-04

PRIOR APPLICATION NUMBER: PCT US97/18008

PRIOR FILING DATE: 1997-10-03

PRIOR APPLICATION NUMBER: 09/286,691

PRIOR FILING DATE: 1999-04-05

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Patentia Ver. 2.0

SEQ ID NO 1

LENGTH: 1558

TYPE: DNA

ORGANISM: Orpinomyces sp. PC-2

FEATURE:

NAME/KEY: CDS

LOCATION: (105)..(1481)

US-09-687-147-1

Query Match 62.6%; Score 19.4; DB 3; Length 1558;

Best Local Similarity 79.3%; Pred. No. 1.1e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TAGGGAACCTTGCTGATATGAAATTT 30

DB 491 TAGGATACCTCTTGGAAATGAAATTT 519

RESULT 6

US-09-539-333D-117/c

Sequence 117, Application US/09539333D

Patent No. 6476208

GENERAL INFORMATION:

APPLICANT: Cohen, Daniel

APPLICANT: Blumenfeld, Marta

APPLICANT: Chumakov, Ilya

APPLICANT: Bougueleret, Lydie

APPLICANT: Bihain, Bernard

APPLICANT: Essioux, Laurent

TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLERIC MARKERS

FILE REFERENCE: GENSET.047AUS

CURRENT APPLICATION NUMBER: US/09/539,333D

CURRENT FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: US 60/126,903

PRIOR FILING DATE: 1999-03-30

PRIOR APPLICATION NUMBER: US 60/131,971

PRIOR FILING DATE: 1999-04-30

PRIOR APPLICATION NUMBER: US 60/132,065

PRIOR FILING DATE: 1999-04-30

PRIOR APPLICATION NUMBER: US 60/143,928

PRIOR FILING DATE: 1999-07-14

PRIOR APPLICATION NUMBER: US 60/145,915

PRIOR FILING DATE: 1999-07-27

PRIOR APPLICATION NUMBER: US 60/146,453

PRIOR FILING DATE: 1999-07-29

PRIOR APPLICATION NUMBER: US 60/146,452

PRIOR FILING DATE: 1999-07-29

PRIOR APPLICATION NUMBER: US 60/162,288

PRIOR FILING DATE: 1999-10-28

PRIOR APPLICATION NUMBER: US 09/416,384

PRIOR FILING DATE: 1999-10-12

NUMBER OF SEQ ID NOS: 231

SOFTWARE: Patent.pm

SEQ ID NO 117

LENGTH: 3001

TYPE: DNA

ORGANISM: Homo Sapiens

FEATURE:

NAME/KEY: allele

LOCATION: 1501

OTHER INFORMATION: 99-15065-85 : polymorphic base C or G

FEATURE:

NAME/KEY: misc binding

LOCATION: 1502..1520

OTHER INFORMATION: 99-15065-85.misl, complement

FEATURE:

NAME/KEY: misc binding

LOCATION: 1481..1500

OTHER INFORMATION: 99-15065-85.mis2,

FEATURE:

NAME/KEY: primer bind

LOCATION: 1568..1585

OTHER INFORMATION: upstream amplification primer, complement

FEATURE:

NAME/KEY: primer bind

LOCATION: 1120..1140

OTHER INFORMATION: downstream amplification primer

FEATURE:

NAME/KEY: misc binding

LOCATION: 1489..1513

OTHER INFORMATION: 99-15065-85 probe

US-09-539-333D-117

Query Match 62.6%; Score 19.4; DB 4; Length 3001;

Best Local Similarity 79.3%; Pred. No. 1.1e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGCGAAGCTTGCTGATATGAAATTT 31

DB 2059 AGCGAAGCTTGCTGATATGAAATTT 2031

RESULT 7

US-09-539-333D-216/c

Sequence 216, Application US/09539333D

Patent No. 6476208

GENERAL INFORMATION:

APPLICANT: Cohen, Daniel

APPLICANT: Blumenfeld, Marta

APPLICANT: Chumakov, Ilya

APPLICANT: Bougueleret, Lydie

APPLICANT: Bihain, Bernard

APPLICANT: Essioux, Laurent

TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLERIC MARKERS

FILE REFERENCE: GENSET.047AUS

CURRENT APPLICATION NUMBER: US/09/539,333D

CURRENT FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: US 60/126,903

PRIOR FILING DATE: 1999-03-30

PRIOR APPLICATION NUMBER: US 60/131,971

PRIOR FILING DATE: 1999-04-30

PRIOR APPLICATION NUMBER: US 60/132,065

PRIOR FILING DATE: 1999-04-30

PRIOR APPLICATION NUMBER: US 60/143,928

PRIOR FILING DATE: 1999-07-14

PRIOR APPLICATION NUMBER: US 60/145,915

PRIOR FILING DATE: 1999-07-27

PRIOR APPLICATION NUMBER: US 60/146,453

```

PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/146,452
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/162,288
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: US 09/416,384
PRIOR FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 231
SOFTWARE: Patent.pm
SEQ ID NO 216
LENGTH: 3001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 1501
OTHER INFORMATION: 99-27297-280 : polymorphic base T or C
FEATURE:
NAME/KEY: misc binding
LOCATION: 1481..1500
OTHER INFORMATION: 99-27297-280.mis1,
FEATURE:
NAME/KEY: misc binding
LOCATION: 1502..1521
OTHER INFORMATION: 99-27297-280.mis2, complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 1761..1779
OTHER INFORMATION: upstream amplification primer, complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 1206..1224
OTHER INFORMATION: downstream amplification primer
FEATURE:
NAME/KEY: misc binding
LOCATION: 1489..1513
OTHER INFORMATION: 99-27297-280 probe
US-09-539-333D-216

Query Match 62.6%; Score 19.4; DB 4; Length 3001;
Best Local Similarity 79.3%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

CY 3 AGTGAAACTGTGTGAATTATGAATT 31
DB 2912 AGTGAAATGTGTGAATTATGAAGTT 2884

RESULT 8
US-09-345-882-1
Sequence 1, Application US/09345882
Patent No. 639373
GENERAL INFORMATION:
APPLICANT: Bouquelerec, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
FILE REFERENCE: GENSET 031A
CURRENT APPLICATION NUMBER: US/09/345,882
CURRENT FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
PRIOR FILING DATE: 1998-12-10
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 162450
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 72794
OTHER INFORMATION: 5-124-273 : polymorphic base A or G

FEATURE:
NAME/KEY: allele
LOCATION: 8073
OTHER INFORMATION: 5-127-261 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:
NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
```



NAME/KEY: allele  
 LOCATION: 150329 : polymorphic base A or G  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 160031  
 OTHER INFORMATION: 5-148-352 : polymorphic base G or T  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 72771..72817  
 OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 72771..72817  
 OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 88050..88096  
 OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 88050..88096  
 OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 90819..90865  
 OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 90819..90865  
 OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 93690..93736  
 OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 93690..93736  
 OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 97099..97145  
 OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 97099..97145  
 OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 97130..97177  
 OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 97130..97177  
 OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 99075..99121  
 OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 99075..99121  
 OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 99094..99140  
 OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 99094..99140  
 OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56  
 FEATURE:  
 NAME/KEY: allele

LOCATION: 103783..103828  
 OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 103783..103828  
 OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 106918..106966  
 OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 106918..106966  
 OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 108084..108130  
 OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 108084..108130  
 OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 108127..108177  
 OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39  
 FEATURE:  
 NAME/KEY: allele  
 LOCATION: 108127..108177  
 OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60  
 FEATURE:

Query Match 61.9%; Score 19.2; DB 4; Length 162450;  
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 TGAACCTGTGTGAATATGAAT 28  
 Db 14062 TGTAACTGTGTGAATATGAAT 14065

RESULT 9  
 US-08-687-080-59  
 Sequence 59, Application US/08687080  
 Patent No. 5965427  
 GENERAL INFORMATION:  
 APPLICANT: Gregory Dolganov  
 TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof  
 NUMBER OF SEQUENCES: 175  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Delinger & Associates  
 STREET: 350 Cambridge Avenue, Suite 250  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94306  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/687,080  
 FILING DATE: 17-JUL-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/592,126  
 FILING DATE: 26-JAN-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sholtz, Charles K.  
 REGISTRATION NUMBER: 38,615  
 REFERENCE/DOCKET NUMBER: 4600-0111.30  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1485 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: 5' END OF INTRON 2 OF RAD50 GENOMIC  
INDIVIDUAL ISOLATE: SEQUENCE  
US-08-687-080-55

Query Match 61.3%; Score 19; DB 2; Length 1485;  
Best Local Similarity 81.5%; Pred. No. 1.4e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 TGAAGCTGTGTATATGAATTT 31  
DB 355 TGAACCTTGTGTATATGAATTT 381

RESULT 10  
PCT-US96-07709-18  
Sequence 18, Application PC/TUS9607709

GENERAL INFORMATION:  
APPLICANT: TIDP, Cynthia A.  
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN  
TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/07709  
FILING DATE: 23-MAY-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-30-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
FAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 594 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..593  
PCT-US96-07709-18

Query Match 60.6%; Score 18.8; DB 5; Length 594;  
Best Local Similarity 76.7%; Pred. No. 1.9e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATAGTGAAGCTGTGTATATGAATTT 30  
DB 12 AATGTGACATCTCTGTATATCAAAAT 41

RESULT 11  
PCT-US96-07709-20/c  
Sequence 20, Application PC/TUS9607709

GENERAL INFORMATION:  
APPLICANT: TIDP, Cynthia A.  
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN  
TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/07709  
FILING DATE: 23-MAY-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-30-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
FAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 594 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US96-07709-20

Query Match 60.6%; Score 18.8; DB 5; Length 594;  
Best Local Similarity 76.7%; Pred. No. 1.9e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATAGTGAAGCTGTGTATATGAATTT 30  
DB 583 AATGTGACATCTCTGTATATCAAAAT 554

RESULT 12  
PCT-US96-07709-27  
Sequence 27, Application PC/TUS9607709

GENERAL INFORMATION:  
APPLICANT: TIDP, Cynthia A.  
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN  
TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/07709  
FILING DATE: 23-MAY-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 744 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US96-07709-27

Query Match 60.6%; Score 18.8; DB 5; Length 744;  
Best Local Similarity 76.7%; Pred. No. 1.9e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

1 ATAGTGAACCTTGCTATATGCAATT 30  
10 AAAGTGACATCTTCTGTATATCAAAAT 39

RESULT 13  
PCT-US96-07709-28/c  
Sequence 28, Application PC/TUS9607709  
GENERAL INFORMATION:  
APPLICANT: TIPP, Cynthia A.  
APPLICANT: Wisniewski, Nancy  
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN  
TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/07709  
FILING DATE: 23-MAY-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-30-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 744 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US96-07709-28

Query Match 60.6%; Score 18.8; DB 5; Length 744;  
Best Local Similarity 76.7%; Pred. No. 1.9e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

1 ATAGTGAACCTTGCTATATGCAATT 30  
12 AAAGTGACATCTTCTGTATATCAAAAT 41

RESULT 14  
PCT-US96-07709-24  
Sequence 24, Application PC/TUS9607709

GENERAL INFORMATION:  
APPLICANT: TIPP, Cynthia A.  
APPLICANT: Wisniewski, Nancy  
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN  
TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/07709  
FILING DATE: 23-MAY-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-30-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 905 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..749

PCT-US96-07709-24

Query Match 60.6%; Score 18.8; DB 5; Length 905;  
Best Local Similarity 76.7%; Pred. No. 1.8e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

1 ATAGTGAACCTTGCTATATGCAATT 30  
12 AAAGTGACATCTTCTGTATATCAAAAT 41

RESULT 15  
PCT-US96-07709-26/c  
Sequence 26, Application PC/TUS9607709

GENERAL INFORMATION:  
APPLICANT: TIPP, Cynthia A.  
APPLICANT: Wisniewski, Nancy  
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN  
TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Sheridan Ross & McIntosh

STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/07709  
FILING DATE: 23-MAY-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gazy J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-30-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 905 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US96-07709-26

Query Match	60.6%;	Score 18.8;	DB 5;	Length 905;
Best Local Similarity	76.7%;	Pred. No. 1.8e+02;		
Matches	23;	Mismatches	7;	Indels 0;
				Gaps 0;

OY            1 ATAGTGAACCTGTGTAATTATGCAATT 30  
| | | | |  
Db         894 AAAGTGACATCTTCGTATTTATCAAAAT 865

Search completed: December 18, 2003, 09:04:20  
Job time : 50.2295 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2003, 09:02:17 ; Search time 998.098 Seconds  
(without alignments)  
103.495 Million cell updates/sec

Title: US-09-889-491-14

Perfect score: 31  
Sequence: 1 atagtgaaactcgtgtaattatgaatttt 31

Scoring table:  
Gapop 10.0, Gapext 1.0

Searched: 221978 seqs, 1666101734 residue

Total number of hits satisfying chosen parameters: 4423956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgm2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgm2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgm2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgm2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgm2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgm2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgm2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgm2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgm2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgm2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 11: /cgm2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq:\*
- 12: /cgm2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgm2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgm2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgm2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgm2\_6/ptodata/2/pubpna/US10C\_NEW\_PUB.seq:\*
- 17: /cgm2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgm2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27.8	89.7	7771	US-10-311-455-1945	Sequence 1945, Ap
2	23	74.2	7771	US-10-311-455-1945	Sequence 1946, Ap
3	22.6	72.9	134	US-09-878-574-8312	Sequence 8312, Ap
4	21.6	69.7	7737	US-09-764-887-453	Sequence 453, App
5	21.6	69.7	7737	US-09-764-887-453	Sequence 2001, Ap
6	21.6	69.7	7737	US-10-092-154-2001	Sequence 2001, Ap
7	21.6	69.7	7737	US-10-073-961-453	Sequence 453, App
8	21.6	69.7	13377	US-10-311-455-1436	Sequence 1436, Ap
9	21.4	69.0	653	US-10-027-632-213332	Sequence 213332, Ap
10	21.4	69.0	653	US-10-027-632-213332	Sequence 213332, Ap
11	21.4	69.0	2704	US-10-106-698-1371	Sequence 1371, Ap
12	21.4	69.0	3747	US-10-071-766-33	Sequence 33, App
13	21.4	69.0	1503841	US-09-795-668-1	Sequence 1, App11
14	21.4	69.0	1503841	US-09-795-668-1	Sequence 1, App11
15	21.4	69.0	1503841	US-09-946-807-1	Sequence 1, App11

C 16	21	67.7	2140405	13	US-10-027-632-76212	Sequence 76212, A
C 17	21	67.7	2140405	13	US-10-027-632-76212	Sequence 1559, Ap
C 18	20.6	66.5	6227	14	US-10-311-455-1559	Sequence 46536, A
C 19	20.4	65.8	473	13	US-10-027-632-46536	Sequence 46536, A
C 20	20.4	65.8	473	13	US-10-027-632-46536	Sequence 195309, A
C 21	20.4	65.8	597	13	US-10-027-632-195309	Sequence 195309, A
C 22	20.4	65.8	597	13	US-10-027-632-195309	Sequence 195309, A
C 23	20.4	65.8	597	14	US-10-027-632-195310	Sequence 195310, A
C 24	20.4	65.8	597	14	US-10-027-632-195310	Sequence 1222, Ap
C 25	20.4	65.8	5163	13	US-10-311-455-1222	Sequence 102215, A
C 26	20.2	65.2	599	13	US-10-027-632-102215	Sequence 102215, A
C 27	20.2	65.2	599	13	US-10-027-632-102215	Sequence 102215, A
C 28	20.2	65.2	599	14	US-10-027-632-102215	Sequence 102215, A
C 29	20.2	65.2	599	14	US-10-027-632-102215	Sequence 206477, A
C 30	20.2	65.2	507	13	US-10-027-632-206477	Sequence 206477, A
C 31	20	64.5	707	14	US-10-027-632-206477	Sequence 149366, A
C 32	20	64.5	740	14	US-10-027-632-149366	Sequence 149366, A
C 33	20	64.5	6682	10	US-09-764-864-1699	Sequence 1699, Ap
C 34	20	64.5	10945	13	US-10-240-453-327	Sequence 227, App
C 35	20	64.5	419	11	US-09-983-985-3215	Sequence 2215, App
C 36	19.8	63.9	424	11	US-09-918-995-37378	Sequence 37378, A
C 37	19.8	63.9	564	13	US-10-027-632-269933	Sequence 269933, A
C 38	19.8	63.9	564	13	US-10-027-632-269933	Sequence 269933, A
C 39	19.8	63.9	564	14	US-10-027-632-269932	Sequence 269932, A
C 40	19.8	63.9	600	13	US-10-027-632-269932	Sequence 269932, A
C 41	19.8	63.9	600	14	US-10-027-632-269932	Sequence 251929, A
C 42	19.8	63.9	1233	13	US-10-027-632-251929	Sequence 251929, A
C 43	19.8	63.9	1233	13	US-10-027-632-251929	Sequence 251929, A
C 44	19.8	63.9	1233	14	US-10-027-632-251930	Sequence 251930, A
C 45	19.8	63.9	1233	14	US-10-027-632-251930	Sequence 251930, A

#### ALIGNMENTS

RESULT 1  
US-10-311-455-1945  
; Sequence 1945, Application US/10311455  
; Publication NO. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLER, Alexander  
; APPLICANT: PIEPERBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detecting Cytosine Methylation  
; TITLE OF INVENTION: Cytosine methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032523.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1945  
; LENGTH: 7771  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)  
US-10-311-455-1945

Query Match 89.7%; Score 27.8; DB 13; Length 7771;  
Best Local Similarity 93.5%; Pred. No. 3;  
Matches 29; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATAGTGAACCTGCTGTAATGAAATTTT 31  
DB 4650 ATAGTGAACCTGCTGTAATGAAATTTT 4680

RESULT 2

US-10-311-455-1946/C  
; Sequence 1946, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OER, Alexander  
; APPLICANT: PIENEBROCK, Christian  
; APPLICANT: BERLIN, Knut  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect  
; TITLE OF INVENTION: Cytosine methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; PRIOR FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1946  
; LENGTH: 7771  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1946

Query Match 74.2%; Score 23; DB 13; Length 7771;  
Best Local Similarity 83.9%; Pred. No. 1.4e+02;  
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATAGTGAACCTGTGTATATGAAATTT 31  
DB 3122 ATATATTAACCTGTGTATATGAAATTT 3092

RESULT 3  
US-09-878-574-8312  
; Sequence 8312, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 8312  
; LENGTH: 134  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 701101005H1  
US-09-878-574-8312

Query Match 72.9%; Score 22.6; DB 10; Length 134;  
Best Local Similarity 86.2%; Pred. No. 1e+02;  
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TAGTGAACCTGTGTATATGAAATTT 30  
DB 39 TAGTGAACCTGTGTATATGAAATTT 67

RESULT 4  
US-09-764-887-453  
; Sequence 453, Application US/09764887  
; Patent No. US20020042096A1  
; GENERAL INFORMATION:

APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P4113  
; CURRENT APPLICATION NUMBER: US/09/764,887  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 658  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 453  
; LENGTH: 7737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-887-453

Query Match 69.7%; Score 21.6; DB 9; Length 7737;  
Best Local Similarity 85.7%; Pred. No. 4.4e+02;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GTGGAACCTGTGTATATGAAATTT 31  
DB 4085 GTGGAACCTGTGTATATGAAATTT 4112

RESULT 5  
US-09-764-847-2001  
; Sequence 2001, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764,847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2001  
; LENGTH: 7737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-847-2001

Query Match 69.7%; Score 21.6; DB 10; Length 7737;  
Best Local Similarity 85.7%; Pred. No. 4.4e+02;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GTGGAACCTGTGTATATGAAATTT 31  
DB 4085 GTGGAACCTGTGTATATGAAATTT 4112

RESULT 6  
US-10-092-154-2001  
; Sequence 2001, Application US/10092154  
; Publication No. US20030054375A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009C1  
; CURRENT APPLICATION NUMBER: US/10/092,154  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2003  
; Prior application removed - See file wrapper or PALM  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2001  
; LENGTH: 7737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-092-154-2001

Query Match 69.7%; Score 21.6; DB 15; Length 7737;  
Best Local Similarity 85.7%; Pred. No. 4.4e+02;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GTGAACCTGTGTAATTCGAAATTT 31  
 Db 4085 GTGAACCTGTGTAATTCGAAATTT 4112

RESULT 7  
 US-10-073-961-453  
 ; Sequence 453, Application US/10073961  
 ; Publication No. US2003007602A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P113C1  
 ; CURRENT APPLICATION NUMBER: US/10/073,961  
 ; CURRENT FILING DATE: 2002-03-05  
 ; PRIOR APPLICATION NUMBER: 09/764,887  
 ; PRIOR FILING DATE: 2001-01-17  
 ; PRIOR APPLICATION NUMBER: 60/179,065  
 ; PRIOR FILING DATE: 2000-01-31  
 ; PRIOR APPLICATION NUMBER: 60/180,628  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: 60/214,886  
 ; PRIOR FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: 60/217,487  
 ; PRIOR FILING DATE: 2000-07-11  
 ; PRIOR APPLICATION NUMBER: 60/225,758  
 ; PRIOR FILING DATE: 2000-08-14  
 ; PRIOR APPLICATION NUMBER: 60/220,963  
 ; PRIOR FILING DATE: 2000-07-26  
 ; PRIOR APPLICATION NUMBER: 60/217,496  
 ; PRIOR FILING DATE: 2000-07-11  
 ; PRIOR APPLICATION NUMBER: 60/225,447  
 ; PRIOR FILING DATE: 2000-08-14  
 ; PRIOR APPLICATION NUMBER: 60/218,290  
 ; PRIOR FILING DATE: 2000-07-14  
 ; PRIOR APPLICATION NUMBER: 60/225,757  
 ; PRIOR FILING DATE: 2000-08-14  
 ; PRIOR APPLICATION NUMBER: 60/226,868  
 ; PRIOR FILING DATE: 2000-08-22  
 ; PRIOR APPLICATION NUMBER: 60/216,647  
 ; PRIOR FILING DATE: 2000-07-07  
 ; PRIOR APPLICATION NUMBER: 60/225,267  
 ; PRIOR FILING DATE: 2000-08-14  
 ; PRIOR APPLICATION NUMBER: 60/216,880  
 ; PRIOR FILING DATE: 2000-07-07  
 ; PRIOR APPLICATION NUMBER: 60/225,270  
 ; PRIOR FILING DATE: 2000-08-14  
 ; PRIOR APPLICATION NUMBER: 60/251,869  
 ; PRIOR FILING DATE: 2000-12-08  
 ; PRIOR APPLICATION NUMBER: 60/235,834  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: 60/234,274  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: 60/234,223  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: 60/228,924  
 ; PRIOR FILING DATE: 2000-08-30  
 ; PRIOR APPLICATION NUMBER: 60/224,518  
 ; PRIOR FILING DATE: 2000-08-14  
 ; PRIOR APPLICATION NUMBER: 60/236,369  
 ; PRIOR FILING DATE: 2000-09-29  
 ; PRIOR APPLICATION NUMBER: 60/224,519  
 ; PRIOR FILING DATE: 2000-08-14  
 ; PRIOR APPLICATION NUMBER: 60/220,964  
 ; PRIOR FILING DATE: 2000-07-26  
 ; PRIOR APPLICATION NUMBER: 60/241,809  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/249,299  
 ; PRIOR FILING DATE: 2000-11-17  
 ; PRIOR APPLICATION NUMBER: 60/236,327  
 ; PRIOR FILING DATE: 2000-09-29  
 ; PRIOR APPLICATION NUMBER: 60/241,785

;; PRIOR FILING DATE: 2000-10-20  
 ;; PRIOR APPLICATION NUMBER: 60/244,617  
 ;; PRIOR FILING DATE: 2000-11-01  
 ;; PRIOR APPLICATION NUMBER: 60/225,268  
 ;; PRIOR FILING DATE: 2000-08-14  
 ;; PRIOR APPLICATION NUMBER: 60/236,368  
 ;; PRIOR FILING DATE: 2000-09-29  
 ;; PRIOR APPLICATION NUMBER: 60/251,856  
 ;; PRIOR FILING DATE: 2000-12-08  
 ;; PRIOR APPLICATION NUMBER: 60/251,868  
 ;; PRIOR FILING DATE: 2000-12-08  
 ;; PRIOR APPLICATION NUMBER: 60/229,344  
 ;; PRIOR FILING DATE: 2000-09-01  
 ;; PRIOR APPLICATION NUMBER: 60/234,997  
 ;; PRIOR FILING DATE: 2000-09-29  
 ;; PRIOR APPLICATION NUMBER: 60/229,343  
 ;; PRIOR FILING DATE: 2000-09-01  
 ;; PRIOR APPLICATION NUMBER: 60/229,345  
 ;; PRIOR FILING DATE: 2000-09-01  
 ;; PRIOR APPLICATION NUMBER: 60/229,287  
 ;; PRIOR FILING DATE: 2000-09-01  
 ;; PRIOR APPLICATION NUMBER: 60/229,513  
 ;; PRIOR FILING DATE: 2000-09-05  
 ;; PRIOR APPLICATION NUMBER: 60/231,413  
 ;; PRIOR FILING DATE: 2000-09-08  
 ;; PRIOR APPLICATION NUMBER: 60/229,509  
 ;; PRIOR FILING DATE: 2000-09-05  
 ;; PRIOR APPLICATION NUMBER: 60/236,367  
 ;; PRIOR FILING DATE: 2000-09-29  
 ;; PRIOR APPLICATION NUMBER: 60/237,039  
 ;; PRIOR FILING DATE: 2000-10-02  
 ;; PRIOR APPLICATION NUMBER: 60/237,038  
 ;; PRIOR FILING DATE: 2000-10-02  
 ;; PRIOR APPLICATION NUMBER: 60/236,370  
 ;; PRIOR FILING DATE: 2000-09-29  
 ;; PRIOR APPLICATION NUMBER: 60/236,802  
 ;; PRIOR FILING DATE: 2000-10-02  
 ;; PRIOR APPLICATION NUMBER: 60/237,037  
 ;; PRIOR FILING DATE: 2000-10-02  
 ;; PRIOR APPLICATION NUMBER: 60/237,040  
 ;; PRIOR FILING DATE: 2000-10-02  
 ;; PRIOR APPLICATION NUMBER: 60/240,960  
 ;; PRIOR FILING DATE: 2000-10-20  
 ;; PRIOR APPLICATION NUMBER: 60/239,935  
 ;; PRIOR FILING DATE: 2000-10-13  
 ;; PRIOR APPLICATION NUMBER: 60/239,937  
 ;; PRIOR FILING DATE: 2000-10-13  
 ;; PRIOR APPLICATION NUMBER: 60/241,787  
 ;; PRIOR FILING DATE: 2000-10-20  
 ;; PRIOR APPLICATION NUMBER: 60/246,474  
 ;; PRIOR FILING DATE: 2000-11-08  
 ;; PRIOR APPLICATION NUMBER: 60/246,532  
 ;; PRIOR FILING DATE: 2000-11-08  
 ;; PRIOR APPLICATION NUMBER: 60/249,216  
 ;; PRIOR FILING DATE: 2000-11-17  
 ;; PRIOR APPLICATION NUMBER: 60/249,210  
 ;; PRIOR FILING DATE: 2000-11-17  
 ;; PRIOR APPLICATION NUMBER: 60/226,681  
 ;; PRIOR FILING DATE: 2000-08-22  
 ;; PRIOR APPLICATION NUMBER: 60/225,759  
 ;; PRIOR FILING DATE: 2000-08-14  
 ;; PRIOR APPLICATION NUMBER: 60/225,213  
 ;; PRIOR FILING DATE: 2000-08-14  
 ;; PRIOR APPLICATION NUMBER: 60/227,182  
 ;; PRIOR FILING DATE: 2000-08-22  
 ;; PRIOR APPLICATION NUMBER: 60/225,214  
 ;; PRIOR FILING DATE: 2000-08-14  
 ;; PRIOR APPLICATION NUMBER: 60/235,836  
 ;; PRIOR FILING DATE: 2000-09-27  
 ;; PRIOR APPLICATION NUMBER: 60/230,438  
 ;; PRIOR FILING DATE: 2000-09-06  
 ;; PRIOR APPLICATION NUMBER: 60/215,135  
 ;; PRIOR FILING DATE: 2000-06-30

```

PRIOR APPLICATION NUMBER: 60/225,266
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/249,218
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,208
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,213
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,212
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,207
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,245
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,244
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,217
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,211
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,215
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,264
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,214
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,297
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/233,400
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/231,242
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,081
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,080
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,414
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,244
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/233,064
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/233,063
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,397
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,399
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,401
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/241,808
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,826
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,786
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,221
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,475
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/231,243
PRIOR FILING DATE: 2000-09-08

```

```

Query Match 69.7%; Score 21.6; DB 15; Length 7737;
Best Local Similarity 85.7%; Pred. No. 4.4e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 4 GTGAACCTTGCTAATTTGAATTT 31
DB 4085 GTGAACCTTGCTAATTTGAATTT 4112

```

RESULT 8

```

US-10-311-455-1436
; Sequence 1436, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OBER, Alexander
; APPLICANT: FISHERBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by De
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1436
; LENGTH: 13377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1436

```

```

Query Match 69.7%; Score 21.6; DB 13; Length 13377;
Best Local Similarity 85.7%; Pred. No. 4.8e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 4 GTGAACCTTGCTAATTTGAATTT 31
DB 1827 GTGAACCTTGCTAATTTGAATTT 1854

```

```

RESULT 9
US-10-027-632-213232
; Sequence 213232, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 213232
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-213232

```

```

Query Match 69.0%; Score 21.4; DB 13; Length 653;
Best Local Similarity 80.6%; Pred. No. 3.4e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```



Qy 1 ATAGTGAAACTGTGTGTAATTAATGAATTTT 31  
Db 486 ATAGTGACATCTGTGTGAATGATCTTAATTTT 516

RESULT 10  
US-10-027-632-213232  
Sequence 213232, Application US/10027632

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: fastseq for Windows Version 4.0

SEQ ID NO 213232

LENGTH: 653

TYPE: DNA

ORGANISM: Human

US-10-027-632-213232

Query Match 69.0%; Score 21.4; DB 14; Length 653;  
Best Local Similarity 80.6%; Pred. No. 3.4e+02;  
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATAGTGAAACTGTGTGTAATTAATGAATTTT 31  
Db 486 ATAGTGACATCTGTGTGAATGATCTTAATTTT 516

RESULT 11  
US-10-106-698-1371  
Sequence 1371, Application US/10106698

GENERAL INFORMATION:

APPLICANT: Ruben et al.

TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

FILE REFERENCE: PA00501

CURRENT APPLICATION NUMBER: US/10/106,698

PRIOR FILING DATE: 2002-03-27

PRIOR APPLICATION NUMBER: PCT/US00/26524

PRIOR FILING DATE: 2000-09-28

PRIOR APPLICATION NUMBER: US 60/157,137

PRIOR FILING DATE: 1999-09-29

PRIOR APPLICATION NUMBER: US 60/163,280

PRIOR FILING DATE: 1999-11-03

NUMBER OF SEQ ID NOS: 8564

SOFTWARE: PatentIn Ver. 3.0

SEQ ID NO 1371

LENGTH: 2704

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: misc feature

LOCATION: (1438)..(1438)

OTHER INFORMATION: n equals a,t,g, or c

US-10-106-698-1371

Query Match 69.0%; Score 21.4; DB 15; Length 2704;  
Best Local Similarity 80.6%; Pred. No. 4.4e+02;  
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATAGTGAAACTGTGTGTAATTAATGAATTTT 31  
Db 966 AAGTTAAATATGTGTATTAATTAATGAATCT 996

RESULT 12  
US-10-071-766-33  
Sequence 33, Application US/10071766

PUBLICATION NO. US20020192678A1

GENERAL INFORMATION:

APPLICANT: Huel-Mei Chen

TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE

FILE REFERENCE: PA-0043 US

CURRENT APPLICATION NUMBER: US/10/071,766

PRIOR FILING DATE: 2002-02-07

NUMBER OF SEQ ID NOS: 144

SOFTWARE: PERL Program

SEQ ID NO 33

LENGTH: 3747

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No. US20020192678A1 1330122.9

LOCATION: 3323-3355

OTHER INFORMATION: a, t, c, g, or other

US-10-071-766-33

Query Match 69.0%; Score 21.4; DB 14; Length 3747;  
Best Local Similarity 80.6%; Pred. No. 4.4e+02;  
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATAGTGAAACTGTGTGTAATTAATGAATTTT 31  
Db 2026 AAGTTAAATATGTGTATTAATTAATGAATCT 2056

RESULT 13  
US-09-795-668-1/c  
Sequence 1, Application US/09795668

PATENT NO. US20020045577A1

GENERAL INFORMATION:

APPLICANT: Stefansson, Hreinn

APPLICANT: Steinhorsdottir, Valgerdur

APPLICANT: Gulcher, Jeffrey R.

TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE

FILE REFERENCE: 2345.2004-001

CURRENT APPLICATION NUMBER: US/09/795,668

PRIOR FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 09/515,716

PRIOR FILING DATE: 2000-02-28

NUMBER OF SEQ ID NOS: 1531

SOFTWARE: fastseq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 1503841

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: misc feature

LOCATION: (1)..(1531)

OTHER INFORMATION: r=g or a

NAME/KEY: misc feature

LOCATION: (1)..(1531)

OTHER INFORMATION: y=t/n or c

NAME/KEY: misc feature

LOCATION: (1)..(1531)

OTHER INFORMATION: m=a or c

```
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c/u
OTHER INFORMATION: k=g or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c
OTHER INFORMATION: s=g or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or t/u
OTHER INFORMATION: w=a or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c or t/u
OTHER INFORMATION: b=g or c or t/u
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or g or t/u
OTHER INFORMATION: d=a or g or t/u
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c or t/u
OTHER INFORMATION: h=a or c or t/u
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or g or c
OTHER INFORMATION: v=a or g or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c or t/u
OTHER INFORMATION: n=a or g or c or t/u
US-09-795-668-1
```

```
Query Match 69.0%; Score 21.4; DB 9; Length 1503841;
Best Local Similarity 80.6%; Pred. No. 9.5e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 ATAGTGAACCTGTGTATATGAAATTTT 31
Db 567651 ATATTGATGCTTGTAATTTTGAATTTT 567621
```

```
RESULT 14
US-09-795-668-1/c
Sequence 1, Application US/09795686
Patent No. US20020094954A1
GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn
APPLICANT: Steinhorsdottir, Valgerdur
APPLICANT: Gulcher, Jeffrey R.
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
FILE REFERENCE: 2345-2005-001
CURRENT APPLICATION NUMBER: US/09/795,686
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 09/515,715
PRIOR FILING DATE: 2000-02-28
NUMBER OF SEQ ID NOS: 1531
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1503841
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or a
OTHER INFORMATION: r=g or a
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c
OTHER INFORMATION: y=c/u or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c
OTHER INFORMATION: m=a or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or t/u
OTHER INFORMATION: k=g or t/u
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c
OTHER INFORMATION: s=g or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c
OTHER INFORMATION: w=a or c or t/u
```

```
OTHER INFORMATION: w=a or t/u
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c or t/u
OTHER INFORMATION: b=g or c or t/u
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c or t/u
OTHER INFORMATION: d=a or g or t/u
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c or t/u
OTHER INFORMATION: h=a or c or t/u
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or g or c
OTHER INFORMATION: v=a or g or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c or t/u
OTHER INFORMATION: n=a or g or c or t/u
US-09-795-668-1
```

```
Query Match 69.0%; Score 21.4; DB 9; Length 1503841;
Best Local Similarity 80.6%; Pred. No. 9.5e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 ATAGTGAACCTGTGTATATGAAATTTT 31
Db 567651 ATATTGATGCTTGTAATTTTGAATTTT 567621
```

```
RESULT 15
US-09-946-807-1/c
Sequence 1, Application US/09946807
Patent No. US20020165144A1
GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn
APPLICANT: Steinhorsdottir, Valgerdur
APPLICANT: Gulcher, Jeffrey R.
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
FILE REFERENCE: 2345-2004-001
CURRENT APPLICATION NUMBER: US/09/946,807
CURRENT FILING DATE: 2001-09-05
PRIOR APPLICATION NUMBER: US/09/795,668
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 09/515,716
PRIOR FILING DATE: 2000-02-28
NUMBER OF SEQ ID NOS: 1531
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1503841
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or a
OTHER INFORMATION: r=g or a
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c
OTHER INFORMATION: y=c/u or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c
OTHER INFORMATION: m=a or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or t/u
OTHER INFORMATION: k=g or t/u
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c
OTHER INFORMATION: s=g or c
NAME/KEY: misc_feature
LOCATION: (1)...(1531) or c
OTHER INFORMATION: w=a or c or t/u
```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1531)
OTHER INFORMATION: b=g or c or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1531)
OTHER INFORMATION: d=a or g or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1531)
OTHER INFORMATION: h=a or c or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1531)
OTHER INFORMATION: v=a or g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1531)
OTHER INFORMATION: n=a or g or c or t/u
US-09-946-807-1
```

```
Query Match      69.0%; Score 21.4; DB 10; length 1503841;
Best Local Similarity 80.6%; Pred. No. 9.5e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy      1 ATAGTGAACCTGTGTGTAATTATGAAATTTT 31
Db      567651 ATATTGATTGCTTGTGTGTAATTGAAATTTT 567621
```

Search completed: December 18, 2003, 10:23:13  
job time : 1014.1 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:02:17 / Search time 1976.38 seconds

(without alignments)  
381.222 Million cell updates/sec

Title: US-09-889-491-14

Perfect score: 31  
Sequence: 1 atagtgaaactgtgtaattgaatttt 31

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_estba:\*  
3: em\_estba:\*  
4: em\_estba:\*  
5: em\_estba:\*  
6: em\_estba:\*  
7: em\_estba:\*  
8: em\_estba:\*  
9: em\_estba:\*  
10: em\_estba:\*  
11: em\_estba:\*  
12: em\_estba:\*  
13: em\_estba:\*  
14: em\_estba:\*  
15: em\_estba:\*  
16: em\_estba:\*  
17: em\_estba:\*  
18: em\_estba:\*  
19: em\_estba:\*  
20: em\_estba:\*  
21: em\_estba:\*  
22: em\_estba:\*  
23: em\_estba:\*  
24: em\_estba:\*  
25: em\_estba:\*  
26: em\_estba:\*  
27: em\_estba:\*  
28: em\_estba:\*  
29: em\_estba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	24.6	79.4	181	CC088689 CSU-K33r.
2	24.6	79.4	242	CC088678 CSU-K33r.
3	24.6	79.4	537	CC115562 NDL_35011
4	24.6	79.4	696	CC074394 CSU-K33r.

Result No.	Score	Query Match Length	ID	Description
5	24.6	79.4	850	CC116481 NDL_60K24
6	24.6	79.4	524	CC088689 CSU-K33r.
7	24.6	79.4	543	CC088689 CSU-K33r.
8	24.6	79.4	555	CC088689 CSU-K33r.
9	24.6	79.4	616	CC088689 CSU-K33r.
10	24.6	79.4	616	CC088689 CSU-K33r.
11	24.6	79.4	814	CC120192 NDL_51F5
12	24.6	79.4	232	CC120192 NDL_51F5
13	24.6	79.4	434	CC120192 NDL_51F5
14	24.6	79.4	553	CC120192 NDL_51F5
15	24.6	79.4	553	CC120192 NDL_51F5
16	24.6	79.4	553	CC120192 NDL_51F5
17	24.6	79.4	553	CC120192 NDL_51F5
18	24.6	79.4	553	CC120192 NDL_51F5
19	24.6	79.4	553	CC120192 NDL_51F5
20	24.6	79.4	553	CC120192 NDL_51F5
21	24.6	79.4	553	CC120192 NDL_51F5
22	24.6	79.4	553	CC120192 NDL_51F5
23	24.6	79.4	553	CC120192 NDL_51F5
24	24.6	79.4	553	CC120192 NDL_51F5
25	24.6	79.4	553	CC120192 NDL_51F5
26	24.6	79.4	553	CC120192 NDL_51F5
27	24.6	79.4	553	CC120192 NDL_51F5
28	24.6	79.4	553	CC120192 NDL_51F5
29	24.6	79.4	553	CC120192 NDL_51F5
30	24.6	79.4	553	CC120192 NDL_51F5
31	24.6	79.4	553	CC120192 NDL_51F5
32	24.6	79.4	553	CC120192 NDL_51F5
33	24.6	79.4	553	CC120192 NDL_51F5
34	24.6	79.4	553	CC120192 NDL_51F5
35	24.6	79.4	553	CC120192 NDL_51F5
36	24.6	79.4	553	CC120192 NDL_51F5
37	24.6	79.4	553	CC120192 NDL_51F5
38	24.6	79.4	553	CC120192 NDL_51F5
39	24.6	79.4	553	CC120192 NDL_51F5
40	24.6	79.4	553	CC120192 NDL_51F5
41	24.6	79.4	553	CC120192 NDL_51F5
42	24.6	79.4	553	CC120192 NDL_51F5
43	24.6	79.4	553	CC120192 NDL_51F5
44	24.6	79.4	553	CC120192 NDL_51F5
45	24.6	79.4	553	CC120192 NDL_51F5

## ALIGNMENTS

RESULT 1  
LOCUS CC088689 181 bp DNA linear GSS 16-APR-2003  
DEFINITION CSU-K33r.34G16.77 CSU-K33r.34G16.77 genomic survey sequence.  
ACCESSION CC088689  
VERSION CC088689.1 GI:29944134  
KEYWORDS GSS.  
SOURCE Aedes aegypti (yellow fever mosquito)  
ORGANISM Aedes aegypti  
REFERENCE 1 (bases 1 to 181)  
AUTHORS Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.  
JOURNAL Unpublished  
COMMENT Contact: Brendan Loftus  
Department of Eukaryotic Genomics  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-3543  
Fax: 301-838-0208  
Email: enta@tigr.org  
Library was provided by Susan Brown and Dennis Knudson at Colorado State University.  
Seq primer: T7

Class: BAC ends.  
Location/Qualifiers  
1. 181  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7159"  
/clone\_1db="CSU-K33r.34G16"  
/note="Vector: pBelobAC11; Site\_1: HindIII"

BASE COUNT 48 a 16 c 51 g 66 t

ORIGIN

Query Match 79.4%; Score 24.6; DB 29; Length 181;  
Best Local Similarity 87.1%; Pred. No. 1.7e+02;  
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATAGTGAAGAACTGTGTGTAATATGAAATTTT 31  
116 ATAGTGAAGAAATTTGAAATATGTAATTTT 86

Db 32 ATAGTGAAGAAATTTGTAATATGTAATTTT 62

RESULT 2  
CC088678/c 242 bp DNA linear GSS 16-APR-2003

LOCUS  
DEFINITION CSU-K33r.34G16.SP6 CSU-K33r Aedes aegypti genomic clone  
ACCESSION CC088678  
VERSION CC088678.1 GI:29944112  
KEYWORDS GSS.  
SOURCE Aedes aegypti (yellow fever mosquito)  
ORGANISM Aedes aegypti  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.  
1 (bases 1 to 242)  
Loftus, B., Shetty, J., Severson, D., Brown, S. and Knudson, D.  
Unpublished  
Other\_GSSs: CSU-K33r.34G16.T7  
Contact: Brendan Loftus  
Department of Eukaryotic Genomics  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-3543  
Fax: 301-838-0208  
Email: entae@tigr.org  
Library was provided by Susan Brown and Dennis Knudson at Colorado State University.  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
Source  
Location/Qualifiers  
1..242  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/strain="Rockville"  
/db\_xref="taxon:7159"  
/clone\_1db="CSU-K33r.34G16"  
/clone\_1db="CSU-K33r"  
/note="Vector: pBelobAC11; Site\_1: HindIII"

BASE COUNT 79 a 64 c 27 g 72 t

ORIGIN

Query Match 79.4%; Score 24.6; DB 29; Length 242;  
Best Local Similarity 87.1%; Pred. No. 1.8e+02;  
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATAGTGAAGAACTGTGTGTAATATGAAATTTT 31  
116 ATAGTGAAGAAATTTGAAATATGTAATTTT 86

Db 116 ATAGTGAAGAAATTTGAAATATGTAATTTT 86

RESULT 3  
CC115562

LOCUS CC115562 537 bp DNA linear GSS 16-APR-2003  
DEFINITION NDL.35011.T7 Notre Dame Liverpool Aedes aegypti genomic clone  
NDL.35011, genomic survey sequence.  
ACCESSION CC115562  
VERSION CC115562.1 GI:29984617  
KEYWORDS GSS.  
SOURCE Aedes aegypti (yellow fever mosquito)  
ORGANISM Aedes aegypti  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.  
1 (bases 1 to 537)  
Loftus, B., Shetty, J., Knudson, D. and Severson, D.  
Unpublished  
Contact: Brendan Loftus  
Department of Eukaryotic Genomics  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-3543  
Fax: 301-838-0208  
Email: entae@tigr.org  
Library was provided by David Severson  
Seq primer: T7  
Class: BAC ends.

FEATURES  
Source  
Location/Qualifiers  
1..537  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/strain="Liverpool"  
/db\_xref="taxon:7159"  
/clone\_1db="NDL.35011"  
/clone\_1db="Notre Dame Liverpool"  
/note="Vector: pBelobAC11; Site\_1: Hind III. The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"

BASE COUNT 147 a 67 c 161 g 162 t

ORIGIN

Query Match 79.4%; Score 24.6; DB 29; Length 537;  
Best Local Similarity 87.1%; Pred. No. 2.1e+02;  
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATAGTGAAGAACTGTGTGTAATATGAAATTTT 31  
40 ATAGTGAAGAAATTTGAAATATGTAATTTT 70

Db 40 ATAGTGAAGAAATTTGAAATATGTAATTTT 70

RESULT 4  
CC074394/c 696 bp DNA linear GSS 16-APR-2003

LOCUS  
DEFINITION CSU-K33r.15020.T7 CSU-K33r Aedes aegypti genomic clone  
ACCESSION CC074394  
VERSION CC074394.1 GI:29915019  
KEYWORDS GSS.  
SOURCE Aedes aegypti (yellow fever mosquito)  
ORGANISM Aedes aegypti  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.  
1 (bases 1 to 696)  
Loftus, B., Shetty, J., Severson, D., Brown, S. and Knudson, D.  
Unpublished  
Other\_GSSs: CSU-K33r.15020.SP6  
Contact: Brendan Loftus  
Department of Eukaryotic Genomics  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-3543  
Fax: 301-838-0208  
Email: entae@tigr.org  
Library was provided by Susan Brown and Dennis Knudson at Colorado State University.

State University.  
Seq primer: T7  
Class: BAC ends.

# FEATURES

Location/Qualifiers  
1..696  
/organism="Aedes aegypti"  
/mol\_type="genomic DNA"  
/strain="Rexville"  
/db\_xref="taxon:7159"  
/clone="CSU-K33r.15020"  
/clone\_lib="CSU-K33r"  
/note=Vector: pBelBAC11, Site\_1: HindIII"

BASE COUNT 199 a 173 c 121 g 203 t

ORIGIN

Query Match 79.4%; Score 24.6; DB 29; Length 696;  
Best Local Similarity 87.1%; Pred. No. 2.2e+02;  
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATAGTGAAACCTGTGTATATGAAATTTT 31  
120 ATAGTGAAACCTGTGTATATGAAATTTT 90

RESULT 5 CC116481 850 bp DNA linear GSS 16-APR-2003

LOCUS ND1.60K24.T7 Notre Dame Liverpool Aedes aegypti genomic clone

DEFINITION ND1.60K24, genomic survey sequence.

ACCESSION CC116481 GI:29985536

VERSION CC116481.1 GI:29985536

KEYWORDS GSS.

SOURCE Aedes aegypti (yellow fever mosquito)

ORGANISM Aedes aegypti

REFERENCE 1 (bases 1 to 850) Knudson, D. and Severson, D.

LOCUS Aedes aegypti

DEFINITION Aedes aegypti

ACCESSION CC116481

VERSION CC116481.1

KEYWORDS GSS.

SOURCE Aedes aegypti (yellow fever mosquito)

ORGANISM Aedes aegypti

REFERENCE 1 (bases 1 to 850) Knudson, D. and Severson, D.

LOCUS Aedes aegypti

DEFINITION Aedes aegypti

ACCESSION CC116481

VERSION CC116481.1

RESULT 6 CNS031MD/c 524 bp DNA linear GSS 01-SEP-2000  
LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone  
DEFINITION 205A06 of library G from Tetradon nigroviridis, genomic survey  
sequence.

ACCESSION AL223726.1 GI:7882559

VERSION AL223726.1

KEYWORDS GSS; genome survey sequence.

SOURCE Tetradon nigroviridis

ORGANISM Tetradon nigroviridis

REFERENCE 1 Roest Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,

AUTHORS Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,

TITLE Saurin, W. and Weissbach, J.

JOURNAL Estimate of human gene number provided by genome-wide analysis

MEDLINE using Tetradon nigroviridis DNA sequence

PUBMED Nat. Genet. 25 (2), 235-238 (2000)

REFERENCE 2 Roest Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,

AUTHORS Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,

TITLE Saurin, W. and Weissbach, J.

JOURNAL Characterization and repeat analysis of the compact genome of the

MEDLINE freshwater pufferfish Tetradon nigroviridis

PUBMED Genome Res. 10 (7), 939-949 (2000)

REFERENCE 3 (bases 1 to 524)

JOURNAL Genoscope.

DEFINITION Direct Submission

LOCUS Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - PRANCE (E-mail : seqrefgenoscope.cns.fr

COMMENT - Web : www.genoscope.cns.fr)

FEATURES This sequence is a single read and was generated as part of a large

source scale clone-end sequencing project of the Tetradon nigroviridis

genome. For more information, please take a look at

http://www.genoscope.cns.fr/Tetradon.

LOCATION/Qualifiers  
1..524  
/organism="Tetradon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="205A06"  
/clone\_lib="G"  
/note=Genoscope sequence ID : CNS0205BA031P1-end : T7"

BASE COUNT 192 a 154 c 0 g 172 t 6 others

ORIGIN

Query Match 74.2%; Score 23; DB 29; Length 524;  
Best Local Similarity 83.9%; Pred. No. 7.5e+02;  
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ATAGTGAAACCTGTGTATATGAAATTTT 31  
89 ATAGTGAAACCTGTGTATATGAAATTTT 59

RESULT 7 AOS15182 543 bp DNA linear GSS 05-MAY-1999

LOCUS HS 2208 B2 E05 T7C CIT Approved Human Genomic Sperm Library D Homo

DEFINITION sapiens genomic clone Plate-2208 Col-10 Row-J, genomic survey

ACCESSION AOS15182

VERSION AOS15182.1 GI:4747440

**KEYWORDS** GSS.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
**REFERENCE** 1 (bases 1 to 543)  
**AUTHORS** Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Kellier, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.  
**TITLE** Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
**MEDLINE** 99380589  
**PUBMED** 10449764  
**COMMENT** Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones may be purchased from Research Genetics (info@resgen.com).  
 BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 2208 Row: J Column: 10  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 543.  
**FEATURES**  
**SOURCE**  
 1. 543  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate=2208 Col=10 Row=J"  
 /sex="male"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /note="Organ: Sperm; Vector: pBelOBAcl1; BAC Clones in E-Coli DH10B"  
**BASE COUNT** 173 a 108 c 99 g 160 t 3 others  
**ORIGIN**  
 Query Match 74.2%; Score 23; DB 28; Length 543;  
 Best Local Similarity 83.9%; Pred. No. 7.5e+02;  
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 1 ATAGTGAAGAACTGTGTATATATGAAATTTT 31  
 447 ATAGTGAAGAACTGTGTATATATGAAATTTT 477  
**Db**  
**RESULT 8** A2997354 595 bp DNA linear GSS 27-APR-2001  
**LOCUS** 2M0283B21R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
**DEFINITION** clone UUGC2M0283B21 R, genomic survey sequence.  
**ACCESSION** A2997354  
**VERSION** A2997354.1 GI:13868581  
**KEYWORDS** GSS.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE** 1 (bases 1 to 595)  
**AUTHORS** Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamll, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, S., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weis, R.  
**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
**JOURNAL** Unpublished  
**COMMENT** Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., STC, UT

84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0283 Row: B Column: 21  
 Seq primer: CACACAGGAAACCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 595.  
**FEATURES**  
**SOURCE**  
 1. 595  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0283B21"  
 /sex="female"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC2M library"  
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/anates/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114[gblAP129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
**BASE COUNT** 155 a 140 c 92 g 207 t 1 others  
**ORIGIN**  
 Query Match 74.2%; Score 23; DB 28; Length 595;  
 Best Local Similarity 83.9%; Pred. No. 7.6e+02;  
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 1 ATAGTGAAGAACTGTGTATATATGAAATTTT 31  
 362 ATAGTGAAGAACTGTGTATATATGAAATTTT 392  
**Db**  
**RESULT 9** A2868713 616 bp DNA linear GSS 21-FEB-2001  
**LOCUS** 2M0180102R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
**DEFINITION** clone UUGC2M0180102 R, genomic survey sequence.  
**ACCESSION** A2868713  
**VERSION** A2868713.1 GI:13072302  
**KEYWORDS** GSS.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE** 1 (bases 1 to 616)  
**AUTHORS** Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamll, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, S., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weis, R.  
**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
**JOURNAL** Unpublished  
**COMMENT** Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: daumgenetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0180 row: 1 column: 02  
 Seq primer: CACACGGAACAGCTATGACC  
 Class: Plasmid ends  
 High quality sequence stop: 616.

## FEATURES

## source

1.616  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="MUGCM0180102"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_id="Mouse 10kb plasmid UGCM library"  
 /note="Vector: PMD42mv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1[473114]pb[AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## BASE COUNT

211 a 76 c 91 g 238 t

## ORIGIN

Query Match 74.2%; Score 23; DB 28; Length 616;  
 Best Local Similarity 83.9%; Pred. No. 7.7e+02;  
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATAGTGAACCTGTGTATATGAAATTTT 31  
 |||||  
 Db 558 ATAGTGAACCTGTGTATATGAAATTTT 588

RESULT 10  
 BQ002494 668 bp mRNA linear EST 26-MAR-2002  
 LOCUS U1-H11-ayw-d-03-0-UI s1 NCI CGAP\_E11 Homo sapiens cDNA clone  
 DEFINITION IMAGE:5844434 3', mRNA sequence.  
 ACCESSION BQ002494  
 VERSION BQ002494.1 GI:19727394  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 668)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strunberg, Ph.D.  
 Email: cgaab@remail.nih.gov  
 Tissue Procurement: Dr. Jose Mercuende  
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/BLMT at: http://image.lml.gov  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

## FEATURES

## source

1.668  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5844434"  
 /tissue\_type="Chondrosarcoma"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_id="NCI CGAP E11"  
 /note="Organ: Left Pelvis; Vector: pRTT3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP E11 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma. The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRTT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (drr)18 tail. The sequence tag for this library is ACCTTGCCAC.  
 TAG LIB=UI-H-E11  
 TAG TISSUE=chondrosarcoma  
 TAG\_SEQ=ACCTTGCCAC"

## BASE COUNT

234 a 96 c 114 g 223 t 1 others

## ORIGIN

Query Match 74.2%; Score 23; DB 12; Length 668;  
 Best Local Similarity 83.9%; Pred. No. 7.8e+02;  
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATAGTGAACCTGTGTATATGAAATTTT 31  
 |||||  
 Db 150 ACAGTGAACCTGTGTATATGAAATTTT 180

RESULT 11  
 CC120192/c 814 bp DNA linear GSS 16-APR-2003  
 LOCUS NDL.51F5.T7 Notre Dame Liverpool Aedes aegypti genomic clone  
 DEFINITION NDL.51F5, genomic survey sequence.  
 ACCESSION CC120192  
 VERSION CC120192.1 GI:29989247  
 KEYWORDS GSS.  
 SOURCE Aedes aegypti (yellow fever mosquito)  
 ORGANISM Aedes aegypti  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicidae; Aedes.  
 REFERENCE 1 (bases 1 to 814)  
 AUTHORS Loftus B., Shetty, J., Koudson, D. and Severson, D.  
 TITLE BAC end sequencing of Aedes aegypti  
 JOURNAL Unpublished  
 COMMENT Contact: Brendan Loftus  
 Department of Bukaryotic Genomics  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-3543  
 Fax: 301-838-0208  
 Email: enta@tigr.org  
 Library was provided by David Severson  
 Seq primer: T7  
 Class: BAC ends.

## FEATURES

## source

1.814  
 /organism="Aedes aegypti"



/mol\_type="genomic DNA"  
/strain="Liverpool"  
/db\_xref="taxon:7159"  
/clone="NDI.51F5"  
/note="Notre Dame Liverpool"  
/note="Vector: pECBAC1, Site 1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"

BASE COUNT 236 a 210 c 149 g 219 t  
ORIGIN

Query Match 74.2%; Score 23; DB 29; length 814;  
Best Local Similarity 83.9%; Pred. No. 8.1e+02;  
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ATAGTGAACCTGTGTATATGAAATTT 31  
|||||  
172 ATAGTGAACCTGTGTATATGAAATTT 142

RESULT 12  
BM85747 232 bp mRNA linear EST 08-MAR-2002  
LOCUS BM85747  
DEFINITION Gm-cl063-4506 5', mRNA sequence.  
ACCESSION Gm-cl063-4506 5', mRNA sequence.  
VERSION BM85747  
KEYWORDS EST.  
SOURCE Glycine max (soybean)  
ORGANISM Glycine max (soybean)

REFERENCE  
AUTHORS Shoemaker, R., Keim, P., Vocklin, L., Expelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, I., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
TITLE Public Soybean EST Project  
JOURNAL Unpublished  
COMMENT Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact: ccu@resgen.com web site: www.resgen.com  
Putative full length read  
vector to vector length is 233  
Seq primer: -40BP from Gibco.

## FEATURES

## SOURCE

1. 232  
/organism="Glycine max"  
/mol\_type="mRNA"  
/db\_xref="taxon:3847"  
/clone="SOYBEAN CLONE ID: Gm-cl063-4506"  
/tissue\_type="germinating shoot, 24 hour germination"  
/lab\_host="DH10B"  
/clone\_lib="Gm-cl063"  
/note="Vector: pBluescript II SK+, Site 1: EcoRI, Site 2: XhoI; The CDNA library was constructed from mRNA isolated from germinating shoots of the cultivar Williams. The seeds were allowed to germinate for 24 hours prior to harvesting the germinating shoots. Complementary DNA was synthesized from mRNA using a primer consisting of a

poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed in the laboratory of Dr. Randy Shoemaker at Iowa State University."

BASE COUNT 106 a 21 c 33 g 72 t  
ORIGIN

Query Match 72.9%; Score 22.6; DB 12; length 232;  
Best Local Similarity 86.2%; Pred. No. 8.7e+02;  
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

2 TAGTGAACCTGTGTATATGAAATTT 30  
|||||  
74 TAGTGAACCTGTGTATATGAAATTT 102

RESULT 13  
BE330627 434 bp mRNA linear EST 04-DEC-2001  
LOCUS BE330627  
DEFINITION Gm-cl040-2149 5', similar to TR:Q43015 Q43015 ALCOHOL DEHYDROGENASE-1CN ;, mRNA sequence.  
ACCESSION BE330627  
VERSION BE330627  
KEYWORDS EST.  
SOURCE Glycine max (soybean)  
ORGANISM Glycine max (soybean)

REFERENCE  
AUTHORS Shoemaker, R., Keim, P., Vocklin, L., Expelding, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, I., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
TITLE Public Soybean EST Project  
JOURNAL Unpublished  
COMMENT Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com  
Insert Length: 618 Std Error: 0.00  
High quality sequence stop: 420.

## FEATURES

## SOURCE

1. 434  
/organism="Glycine max"  
/mol\_type="mRNA"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl040-2149"  
/tissue\_type="Hypocotyl and Plumule, germinating seeds"  
/lab\_host="DH10B"  
/clone\_lib="Gm-cl040"  
/note="Vector: pT73Pac (Pharmacia); Site 1: EcoRI; Site 2: NotI; This cDNA library was constructed from mRNA isolated from hypocotyl and plumule tissues of seeds germinated for three days of the cultivar Williams 82. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a NotI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by digestions with EcoRI and NotI. The

```

/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl087-5369"
/tissue_type="roots"
/lab_host="DH10B"
/clone_lib="Gm-cl087"
/notes="vector: pbluescript II SK+ site.1: EcoRI; site.2:
XhoI. The mRNA was prepared using polyattract mRNA system
from PROMEGA. The cDNA was prepared using the STRATAGENE
kit. Complementary DNA was synthesized from mRNA using a
primer consisting of a poly(dT) sequence with a XhoI
restriction site
(5'GAGAGACAGAGAGAGAGAGACACACTCTCCAGCTTTTTTTTTTTTTTT).
EcoRI adapters (5'-OH-AATTCGACCAAG and 3'-GGCGTCTCT) were

```

Search completed: December 18, 2003, 11:35:09  
Job time : 1984.38 secs

Search completed: December 18, 2003, 11:35:09  
Job time : 1984.38 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2003, 09:02:17 ; Search time 196.721 seconds  
(without alignments)  
411.664 Million cell updates/sec

Title: US-09-889-491-13

Perfect score: 30

Sequence: 1 atagagaagcccaagaataacgctgacc 30

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Genseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	30	21	AAAT4857
2	30	100.0	2421	21	AAAT4857
3	22.2	74.0	2577	23	AAAT4857
4	22.2	74.0	17087	20	AAAT4857
5	22.2	74.0	17087	24	AAAT4857
6	20.4	68.0	2528	24	AAAT4857
7	20.4	68.0	6816	22	AAAT4857
8	20	66.7	3610	24	AAAT4857

Result No.	Score	Query Match	Length	DB ID	Description
9	20	66.7	3610	24	AAAT4857
10	20	66.7	3610	24	AAAT4857
11	20	66.7	5088	23	AAAT4857
12	20	66.7	2365589	24	AAAT4857
13	19.6	65.3	508	22	AAAT4857
14	19.6	65.3	547	22	AAAT4857
15	19.6	65.3	817	22	AAAT4857
16	19.6	65.3	1103	22	AAAT4857
17	19.6	65.3	1786	22	AAAT4857
18	19.6	65.3	52354	24	AAAT4857
19	19.6	65.3	406	23	AAAT4857
20	19.4	64.7	460	22	AAAT4857
21	19.4	64.7	623	22	AAAT4857
22	19.4	64.7	623	22	AAAT4857
23	19.4	64.7	623	22	AAAT4857
24	19.4	64.7	623	22	AAAT4857
25	19.4	64.7	902	24	AAAT4857
26	19.4	64.7	2507	22	AAAT4857
27	19.4	64.7	6854	16	AAAT4857
28	19.4	64.7	6854	19	AAAT4857
29	19.4	64.7	6854	20	AAAT4857
30	19.2	64.0	1695	24	AAAT4857
31	19.2	64.0	2169	23	AAAT4857
32	19.2	64.0	2478	13	AAAT4857
33	19.2	64.0	3466	16	AAAT4857
34	19.2	64.0	20561	20	AAAT4857
35	19.2	64.0	20561	24	AAAT4857
36	19.2	64.0	20561	24	AAAT4857
37	19.2	64.0	483	22	AAAT4857
38	19.2	64.0	483	22	AAAT4857
39	19.2	64.0	483	22	AAAT4857
40	19.2	64.0	483	22	AAAT4857
41	19.2	64.0	483	22	AAAT4857
42	19.2	64.0	483	22	AAAT4857
43	19.2	64.0	483	22	AAAT4857
44	19.2	64.0	483	22	AAAT4857
45	19.2	64.0	483	22	AAAT4857

#### ALIGNMENTS

RESULT 1

ID AAA74857 standard; DNA; 30 BP.

AC AAA74857;

XX 17-JAN-2001 (first entry)

DE Human bone sialoprotein gene promoter BSP-A1496G polymorphic site.

KW Human; bone sialoprotein; BSP; bone mineral density; atherosclerosis;

OS Homo sapiens.

XX Key Location/Qualifiers

XX variation /tag= a

XX /note= "BSP-A1496G polymorphism"

XX W0200042216-A2.

XX 20-UTL-2000.

XX 17-JAN-2000; 2000WO-EP00319.

XX 18-JAN-1999; 99GB-0001037.

XX 28-MAY-1999; 99GB-0012585.

XX (OSTR-) OSTROMETER BIOTECH AS.

PI Kusk P;  
 XX WPI; 2000-476070/41.  
 XX Assessing predisposition to a calcification condition status such as  
 PT osteoporosis and atherosclerosis, comprising determining the genotype  
 PT of the promoter of the bone sialoprotein, matrix gla protein,  
 PT osteopontin or osteoprotegerin gene -  
 XX Dielosure; Fig 1; 70pp; English.  
 XX The present sequence shows the polymorphic site at position 1496 of the  
 CC human bone sialoprotein (BSP) gene promoter. It was used in the methods  
 CC of the invention, which involve the determination of the promoter  
 CC sequence at polymorphisms in a number of genes controlling bone mineral  
 CC density. These genes include sequences encoding the bone sialoprotein,  
 CC the matrix gla protein (MGP), osteopontin (OPN) and  
 CC osteoprotegerin/osteoclastogenesis inhibitory factor (OPG/OCIF). G at  
 CC this position is associated with a higher peak bone mass. Along with the  
 CC other polymorphic sites, this variable sequence can be used to predict an  
 CC individual's predisposition to osteoporosis and atherosclerosis, thus  
 CC enabling earlier treatment and preventive measures.  
 XX Sequence 30 BP; 14 A; 7 C; 5 G; 4 T; 0 other;  
 SO Query Match 100.0%; Score 30; DB 21; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 0.0089;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ATATGAGAGCCCAAGAAATCAGCTGACC 30  
 Db 1 ATATGAGAGCCCAAGAAATCAGCTGACC 30  
 RESULT 2  
 ID AAG61936 standard; DNA; 2421 BP.  
 AC AAG61936;  
 XX 20-NOV-2000 (first entry)  
 DT Human bone sialoprotein (BSP) promoter (-2184-+237).  
 DE  
 XX Bone sialoprotein promoter; BSP; human; tissue-specific promoter;  
 KW targeted expression; therapeutic gene; gene therapy;  
 KW osteotropic tumour; calcification; prostate tumour; osteosarcoma;  
 KW metastatic tumour; lung; breast; colon; brain; multiple myeloma;  
 KW benign prostatic hypertrophy; BPH; arteriosclerosis; osteogenesis;  
 KW osteoblast; bone repair; reporter construct; chromosome 4; ds.  
 XX Homo sapiens.  
 OS WO200036919-A1.  
 PN 29-JUN-2000.  
 PD 22-DEC-1999; 99WO-US30642.  
 PF 22-DEC-1998; 98US-0113200.  
 PR 22-DEC-1998; 98US-0113200.  
 XX (UYVI-) UNIV VIRGINIA PATENT FOUND.  
 PA  
 XX Koehnen KS, Chung LMK;  
 PI WPI; 2000-442489/38.  
 DR Novel therapeutic agent for prevention and treatment of calcified  
 PT tumors and proliferative disorders such as osteosarcoma, multiple  
 PT myeloma and breast cancer, comprises bone sialoprotein promoter -  
 XX Claim 5; Fig 8A-B; 79pp; English.  
 XX

CC The invention relates to a novel gene therapy composition comprising a  
 CC mammalian bone sialoprotein (BSP) promoter operably linked to a gene  
 CC sequence encoding a toxic and/or therapeutic protein contained within a  
 CC delivery vector, such as a recombinant adenovirus or a liposome. The BSP  
 CC promoter directs gene expression in tissue and tumour cells with  
 CC calcification potential. It is active in fully-differentiated osteoblasts  
 CC in bone, in tropoblastic cells of the placenta, and in cementum and  
 CC dentin of teeth. It is also active in a variety of human tumours or  
 CC non-malignant growths which exhibit the ability to calcify either in  
 CC the primary or at metastatic sites. This is especially evident for  
 CC prostate and osteosarcoma tumours, but includes any osteotropic  
 CC aggressive metastatic tumour such as, for example, lung, multiple  
 CC myeloma, breast, colon and brain. The BSP promoter may also be expressed  
 CC in non-malignant conditions in which calcification occurs, such as benign  
 CC prostatic hypertrophy (BPH) and arteriosclerosis. The gene therapy  
 CC composition of the invention may be used to treat osteotropic tumours  
 CC (or non-malignant growths with calcification potential) via delivery and  
 CC expression of a gene encoding a protein such as herpes simplex virus  
 CC thymidine kinase (HSVTK), which converts prodrugs such as acyclovir and  
 CC ganciclovir to a cytotoxic phosphorylated form. The composition may also  
 CC be used to promote bone repair via delivery and expression of a gene  
 CC encoding a therapeutic protein such as a growth factor, a cytokine or  
 CC inhibitor thereof, a differentiation factor, a chemokine, an interferon,  
 CC a colony stimulating factor, or an angiogenic factor. The invention  
 CC also discloses a composition in which the gene under the control of  
 CC the BSP promoter is a reporter gene. Such reporter compositions may  
 CC be used to identify compounds capable of modulating osteotropic-specific  
 CC gene expression. The present sequence represents a human BSP promoter  
 CC (-2184-+237) which is claimed for use in the compositions of the  
 CC invention. The human BSP gene is located on chromosome 4.  
 XX Sequence 2421 BP; 891 A; 438 C; 425 G; 667 T; 0 other;  
 SO Query Match 100.0%; Score 30; DB 21; Length 2421;  
 Best Local Similarity 100.0%; Pred. No. 0.015;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ATATGAGAGCCCAAGAAATCAGCTGACC 30  
 Db 1497 ATATGAGAGCCCAAGAAATCAGCTGACC 1516  
 RESULT 3  
 ID AAS53160 standard; DNA; 2577 BP.  
 AC AAS53160;  
 XX 13-FEB-2002 (first entry)  
 DT Enterococcus faecalis DNA for cellular proliferation protein #588.  
 DE  
 XX Antisense; ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX Enterococcus faecalis.  
 OS WO200170955-A2.  
 PN 27-SEP-2001.  
 PD 21-MAR-2001; 2001WO-US09180.  
 PF 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 23-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI; 2001-611495/70.  
XX P-PSDB; A035301.  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
PS  
PS  
PS Claim 27; Seq ID No 6797; 51bp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *aerogenes*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2577 BP; 861 A; 458 C; 574 G; 684 T; 0 other;  
Query Match 74.0%; Score 22.2; DB 23; Length 2577;  
Best Local Similarity 88.9%; Pred. No. 21;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0  
CY 4 TAGAAGCCCAAGAAAAATCAGCTGAC 30  
DB 1520 TAGAAGCCGAAAGAAAAATCAGTTGAC 1546  
|||||  
RESULT 4  
AA012984/C  
ID AA012984 standard; DNA; 17087 BP.  
XX  
XX AA012984;  
AC  
XX  
XX 19-MAR-1999 (first entry)  
DT  
XX  
XX Enterococcus faecalis genome contig SEQ ID NO:47.  
DE  
XX  
XX Enterococcus faecalis; contig; detection; Enterococcal infection;  
KM vaccine; attenuation; computer readable medium; ds.  
XX  
XX Enterococcus faecalis.  
OS  
XX  
XX WO9850555-A2.  
PN  
XX  
XX 12-NOV-1998.  
PD  
XX  
XX 04-MAY-1998; 98WO-US08985.  
PF  
XX  
XX 14-NOV-1997; 97US-0066009.  
PR  
XX  
XX 06-MAY-1997; 97US-0044031.  
PR  
XX  
XX 16-MAY-1997; 97US-0046655.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Barash SC, Dillon PJ, Kunsch CA;  
PI  
XX  
XX WPI; 1999-045171/04.  
DR

PT New isolated *Enterococcus faecalis* polymucleotides and polypeptides  
 PT - used to develop products for the detection of *Enterococcus* and for  
 PT use in vaccines for prevention or attenuation of *Enterococcus*  
 PT infection.  
 XX  
 PS Claim 1; Page 429-438; 2084dp; English.  
 XX  
 CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the *Enterococcus faecalis* genome.  
 CC AA11238 to AA11319 represent these nucleotide sequences which are  
 CC system nucleotide sequences, also known as contigs. The computer-based  
 CC program can identify fragments of the *Enterococcus faecalis* genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of *Enterococcus faecalis* in samples. They can also be used for  
 CC diagnosing *Enterococcal* infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of *Enterococcus faecalis*, or  
 CC other related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an *Enterococcal*  
 CC infection.  
 XX  
 SQ Sequence 17087 BP; 4734 A; 3683 C; 3245 G; 5415 T; 10 other;  
 QY  
 Query Match 74.0%; Score 22.2; DB 20; Length 17087;  
 Best Local Similarity 88.9%; Pred. No. 26;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps  
 4 TAGAAGCCCAAGAAATATCGCTGACC 30  
 11327 TAGAAGCCGAAGAAATATCGCTGACC 11301  
 RESULT 5  
 ABS98779/C  
 ID ABS98779 standard; DNA; 17087 BP.  
 XX  
 XX ABS98779;  
 XX  
 DT 18-DEC-2002 (first entry)  
 XX  
 DE *Enterococcus faecalis* contig sequence #47.  
 XX  
 XX Computer readable medium; *Enterococcus faecalis*; microbe; growth;  
 KW pathogenicity; vaccine; resistance; *Enterococcal* infection; commercial;  
 KW therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;  
 KW biotech technology; antibacterial; modulator of nucleic acid expression;  
 XX contig; ds.  
 XX  
 OS *Enterococcus faecalis*.  
 XX  
 PN US2002120116-A1.  
 XX  
 PD 29-AUG-2002.  
 XX  
 PF 04-MAY-1998; 98US-0070927.  
 XX  
 PR 04-MAY-1998; 98US-0070927.  
 XX  
 PA (KUNS/) KUNSCH C A.  
 PA (DILL/) DILLON P J.  
 PA (BARA/) BARASH S.  
 XX  
 PI Kunech CA, Dillon PJ, Barash S;  
 XX  
 DR WPI; 2002-750065/81.  
 XX  
 PT Computer readable medium having recorded on it a *Enterococcus faecalis*  
 PT nucleotide sequence useful for detecting diseases related to  
 PT *Enterococcus* infections in animals -  
 PS Claim 1; Page -; 119pp; English.  
 XX

CC The present invention relates to a new computer readable medium with an  
 CC Enterococcus faecalis nucleotide sequence. The invention is useful to  
 CC diagnose the presence of *E. faecalis* in a sample or determining the  
 CC presence of a specific microbe in a sample. The invention is also useful  
 CC for modulating the growth or pathogenicity of *E. faecalis*, in a vaccine  
 CC to confer resistance to Enterococcal infection, for commercial,  
 CC therapeutic and industrial purposes, and for fermenting a particular  
 CC sugar source or to produce a particular metabolite. The invention is  
 CC useful for detecting diseases related to Enterococcus infections in  
 CC animals, and for detecting *E. faecalis* using biotech technology. The  
 CC present nucleic acid sequence represents an Enterococcus faecalis contig  
 CC DNA sequence of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification but was obtained in electronic format directly  
 CC from USPTO at <http://seqdata.uspto.gov>.  
 XX

SO Sequence 17087 BP; 4734 A; 3683 C; 3245 G; 5415 T; 10 other;

Query Match 74.0%; Score 22.2; DB 24; Length 17087;  
 Best Local Similarity 88.9%; Pred. No. 26;  
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TAGAAGCCAGAGAAATCAGCTGACC 30  
 DB 11327 TAGAAGCCAGAGAAATCAGCTGACC 11301

RESULT 6  
 ID ABA05477 standard; cDNA; 2528 BP.  
 AC ABA05477;  
 DT 01-MAR-2002 (first entry)  
 DE Human RNA gyrase 12 encoding cDNA.  
 XX  
 XX Human; RNA gyrase 12; malignant tumour; haemopathy; HIV; infection;  
 KW human immunodeficiency virus; immunological disease; inflammation;  
 KW enzyme; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key location/Qualifiers  
 FT CDS 467..802  
 FT /\*tag= a  
 FT /product= "RNA gyrase 12"  
 FT /note= "claimed in claim 6"  
 XX  
 XX CN1311325-A.  
 PD 05-SEP-2001.  
 PF 02-MAR-2000; 2000CN-0111862.  
 PR 02-MAR-2000; 2000CN-0111862.  
 PA (BODE-) BODE GENB DEV CO LTD SHANGHAI.  
 PI Mao Y, Xie Y;  
 DR WPI; 2002-049927/07.  
 DR P-PSDB; AAM47948.  
 XX  
 XX New polypeptide-human RNA unwindase 12 and polynucleotide encoding the  
 PT polypeptide -  
 PS Claim 6; Page 25-26 (Disclosure); 33pp; Chinese.  
 CC The invention relates to human RNA gyrase 12, the polynucleotide encoding  
 CC this polypeptide and DNA recombinant techniques to produce this  
 CC polypeptide. The present invention also discloses a method of applying  
 CC this polypeptide to treat various diseases, such as malignant tumour,

CC haemopathy, HIV infection, immunological diseases and various  
 CC inflammations.  
 XX

SO Sequence 2528 BP; 815 A; 518 C; 568 G; 627 T; 0 other;

Query Match 68.0%; Score 20.4; DB 24; Length 2528;  
 Best Local Similarity 80.0%; Pred. No. 11e+02;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATATAGAGCCAGAGAAATCAGCTGACC 30  
 DB 2155 ATATAGAGCCAGAGAAATCAGCTGACC 2164

RESULT 7  
 ID AAS4668/c  
 AC AAS4668;  
 DT 18-DEC-2001 (first entry)  
 DE Tumour suppressor gene derived chemically modified sequence #411.  
 XX  
 XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200168912-A2.  
 PD 20-SEP-2001.  
 PF 15-MAR-2001; 2001WO-EP02955.  
 PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1018058.  
 PR 07-APR-2000; 2000DE-1013173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2001-602752/68.  
 PT Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for  
 PT analysing diseases associated with cytosine methylation state e.g.  
 PT cancer -  
 XX  
 PS Claim 1; SEQ ID No 411; 27pp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a  
 CC peptide nucleic acid-oligonucleotide (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with CpG dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC and/or therapy of existing diseases or the predisposition to specific  
 CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and

CC ONCogenes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 6816 BP; 1903 A; 67 C; 1504 G; 338 T; 4 other;

Query Match 68.0%; Score 20.4; DB 22; Length 6816;  
Best Local Similarity 80.0%; Pred. No. 1.2e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATATAGAGCCCAAGAAATCAGCTGACC 30  
Db 6609 ATATAGAGCCCAAGAAATCAGCTGACC 6580

RESULT 8  
ABN95909  
ID ABN95909 standard; DNA; 3610 BP.

XX  
AC ABN95909;

DT 13-AUG-2002 (first entry)

DE Gene #2407 used to diagnose liver cancer.

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KM metastatic liver tumour; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.

XX Homo sapiens.

XX WO200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US30589.

XX 02-OCT-2000; 2000US-237054P.

XX (GENE-) GENE LOGIC INC.

XX Horne D, Alvares C, Perea-da-Silva S, Vockley JG;

XX WPI; 2002-426119/45.

PT Diagnosing and detecting the progression of liver cancer,  
PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
PT involves detecting the level of expression of two or more genes in a  
PT liver tissue sample -

XX Claim 1; SEQ ID NO 2407; 2988bp; English.

XX The invention relates to a novel method for diagnosing and detecting the  
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver  
XX tumor in a patient, and differentiating metastatic liver cancer from  
XX hepatocellular carcinoma in a patient, involving detecting the level of  
XX expression of two or more genes represented in ABN93503-ABN97455 in a  
XX tissue sample. The method of the invention has hepatotropic, and  
XX cytostatic activity. The method is useful for diagnosing and detecting  
XX the progression of liver cancer, hepatocellular carcinoma and metastatic  
XX liver carcinoma in a patient. The method is useful for identifying  
XX expression profiles which serve as useful diagnostic markers as well as  
XX markers that can be used to monitor disease states, disease progression,  
XX drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3610 BP; 1322 A; 622 C; 963 G; 703 T; 0 other;

Query Match 66.7%; Score 20; DB 24; Length 3610;

Best Local Similarity 82.1%; Pred. No. 1.6e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 ATAGAGCCCAAGAAATCAGCTGACC 30  
Db 1112 ATAGAGCCCAAGAAATCAGCTGACC 1139

RESULT 9  
ABL62384  
ID ABL62384 standard; DNA; 3610 BP.

XX  
AC ABL62384;

DT 15-MAY-2002 (first entry)

DE Colon adenocarcinoma related gene sequence SEQ ID NO:721.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KM stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;  
XX gene; ds.

XX Homo sapiens.

XX WO200194629-A2.

XX 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US10838.

XX 05-JUN-2000; 2000US-209473P.

XX 18-SEP-2000; 2000US-209531P.

XX 18-SEP-2000; 2000US-233133P.

XX 20-SEP-2000; 2000US-233617P.

XX 20-SEP-2000; 2000US-234009P.

XX 20-SEP-2000; 2000US-234034P.

XX 22-SEP-2000; 2000US-234052P.

XX 22-SEP-2000; 2000US-234509P.

XX 22-SEP-2000; 2000US-234567P.

XX 25-SEP-2000; 2000US-234923P.

XX 25-SEP-2000; 2000US-234924P.

XX 25-SEP-2000; 2000US-235077P.

XX 25-SEP-2000; 2000US-235082P.

XX 25-SEP-2000; 2000US-235134P.

XX 25-SEP-2000; 2000US-235280P.

XX 26-SEP-2000; 2000US-235637P.

XX 26-SEP-2000; 2000US-235638P.

XX 27-SEP-2000; 2000US-235711P.

XX 27-SEP-2000; 2000US-235720P.

XX 27-SEP-2000; 2000US-235840P.

XX 27-SEP-2000; 2000US-235863P.

XX 28-SEP-2000; 2000US-236028P.

XX 28-SEP-2000; 2000US-236032P.

XX 28-SEP-2000; 2000US-236033P.

XX 28-SEP-2000; 2000US-236034P.

XX 28-SEP-2000; 2000US-236109P.

XX 28-SEP-2000; 2000US-236111P.

XX 28-SEP-2000; 2000US-236842P.

XX 28-SEP-2000; 2000US-236851P.

XX 02-OCT-2000; 2000US-237112P.

XX 02-OCT-2000; 2000US-237173P.

XX 02-OCT-2000; 2000US-237278P.

XX 02-OCT-2000; 2000US-237284P.

XX 02-OCT-2000; 2000US-237295P.

XX 02-OCT-2000; 2000US-237316P.

XX 03-OCT-2000; 2000US-237425P.

XX 03-OCT-2000; 2000US-237598P.

XX 03-OCT-2000; 2000US-237604P.

XX 03-OCT-2000; 2000US-237606P.

XX 03-OCT-2000; 2000US-237608P.

XX 01-NOV-2000; 2000US-244867P.

XX 01-NOV-2000; 2000US-245084P.







PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Deeds J, Berger A, Zhao X;  
 XX  
 DR WPI, 2001-375006/39.  
 XX  
 PT New isolated nucleic acid for diagnosing and treating cervical cancer  
 XX and for assessing and detecting compounds for treating the cancer -  
 XX  
 PS Claim 1; Page 670; 1051pp; English.  
 XX  
 The invention relates to novel genes (AAH68727-AAH73383) associated with  
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
 CC polypeptides are useful to assess if a patient is afflicted with  
 CC cervical cancer or has a pre-malignant condition; to monitor the  
 CC progression of cervical cancer or a premalignant condition in a patient;  
 CC and to select and/or assess the efficacy of a compound or therapy for  
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
 CC useful for gene therapy.  
 XX  
 SQ Sequence 508 BP; 108 A; 143 C; 129 G; 128 T; 0 other;  
 XX  
 Query Match 65.3%; Score 19.6; DB 22; Length 508;  
 Best Local Similarity 84.6%; Pred. No. 1.9e+02;  
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 XX  
 QY 2 TATGAGCCGCAAGAAATCAGCTG 27  
 Db 144 TAGAGAGCCCAAGAGTAGAGCTG 119  
 |||||  
 RESULT 14  
 ID AAH69370/c  
 AC AAH69370 standard; cDNA; 539 BP.  
 XX  
 DT 19-SEP-2001 (first entry)  
 XX  
 DE Human cervical cancer marker nucleic acid 644.  
 XX  
 KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200142467-A2.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PE 08-DEC-2000; 2000WO-US33312.  
 XX  
 PR 08-DEC-1999; 99US-0169681.  
 XX  
 PR 21-DEC-1999; 99US-0171350.  
 XX  
 PR 14-MAR-2000; 2000US-0189315.  
 XX  
 PR 12-MAY-2000; 2000US-0203791.  
 XX  
 PR 09-JUN-2000; 2000US-0210600.  
 XX  
 PR 21-JUN-2000; 2000US-0220114.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Deeds J, Berger A, Zhao X;  
 XX  
 DR WPI, 2001-375006/39.  
 XX  
 PT New isolated nucleic acid for diagnosing and treating cervical cancer  
 XX and for assessing and detecting compounds for treating the cancer -  
 XX  
 PS Claim 1; Page 215-216; 1051pp; English.  
 XX  
 The invention relates to novel genes (AAH68727-AAH73383) associated with  
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
 CC polypeptides are useful to assess if a patient is afflicted with

CC	cervical cancer or has a pre-malignant condition; to monitor the progression of cervical cancer or a premalignant condition in a patient;
CC	to select and/or assess the efficacy of a compound or therapy for inhibiting cervical cancer in a patient. The nucleic acids may also be useful for gene therapy.
XX	
SQ	Sequence 539 BP; 118 A; 152 C; 137 G; 132 T; 0 other;
Query Match	65.3%; Score 19.6; DB 22; Length 539;
Best Local Similarity	84.6%; Pred. No. 1.9e+02;
Matches	22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY	2 TATAGAAAGCCCCAAGAAAATCAGCTG 27
Dd	152 TAGGAAGCCCAAGAGTAGTACACTG 127
RESULT 15	
ID	AAH71172/c
XX	AAH71172 standard; cDNA; 547 BP.
AC	AAH71172;
DT	19-SEP-2001 (first entry)
XX	
DE	Human cervical cancer marker nucleic acid 2446.
KM	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
OS	Homo sapiens.
XX	
PN	WO200142467-A2.
PD	14-JUN-2001.
XX	
PR	08-DEC-2000; 2000MO-US33312.
XX	
PR	08-DEC-1999; 99US-0169681.
PR	21-DEC-1999; 99US-0171350.
PR	14-MAR-2000; 2000US-0189315.
PR	12-MAY-2000; 2000US-0203791.
PR	09-JUN-2000; 2000US-0210600.
PR	21-JUN-2000; 2000US-0220114.
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	
P1	Schlegel R, Deeds J, Berger A, Zhao X;
PI	WPI; 2001-375006/39.
DR	
XX	
PT	New isolated nucleic acid for diagnosing and treating cervical cancer
FT	and for assessing and detecting compounds for treating the cancer -
XX	
PS	Claim 1; Page 513; 105ipp; English.
XX	
CC	The invention relates to novel genes (AAH68727-AAH73383) associated with
CC	cervical cancer with cytotstatic activity. The nucleic acids and encoded
CC	polypeptides are useful: to assess if a patient is afflicted with
CC	cervical cancer or has a pre-malignant condition; to monitor the
CC	progression of cervical cancer or a premalignant condition in a patient;
CC	and to select and/or assess the efficacy of a compound or therapy for
CC	inhibiting cervical cancer in a patient. The nucleic acids may also be
CC	useful for gene therapy.
XX	
SQ	Sequence 547 BP; 119 A; 153 C; 142 G; 133 T; 0 other;
Query Match	65.3%; Score 19.6; DB 22; Length 547;
Best Local Similarity	84.6%; Pred. No. 1.9e+02;
Matches	22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY	2 TATAGAAAGCCCAGAATAATCAGCTG 27
Dd	161 TAGGAAGCCCAAGAGTAGTACACTG 136

Thu Dec 18 13:30:45 2003

us-09-889-491-13.rng

Page 9

Search completed: December 18, 2003, 10:29:56  
Job time : 203.721 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2003, 09:02:17 ; Search time 43.7705 Seconds  
(without alignments)  
302.521 Million cell updates/sec

Title: US-09-889-491-13  
Perfect score: 30  
Sequence: 1 atataagaagcccaagaagaatcagctgacc 30

Scoring table:  
IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgm2\_6/ptodata/2/ina/5A COMB.seg.\*  
2: /cgm2\_6/ptodata/2/ina/5B COMB.seg.\*  
3: /cgm2\_6/ptodata/2/ina/6A COMB.seg.\*  
4: /cgm2\_6/ptodata/2/ina/6B COMB.seg.\*  
5: /cgm2\_6/ptodata/2/ina/PTUS COMB.seg.\*  
6: /cgm2\_6/ptodata/2/ina/backfile1.seg.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	19.4	64.7	US-08-468-036-4	Sequence 4, Appli
2	19.4	64.7	US-08-376-843-4	Sequence 4, Appli
3	19.2	64.0	US-08-441-139-17	Sequence 17, Appli
4	19.2	64.0	US-08-468-036-38	Sequence 38, Appli
5	19.2	64.0	US-08-376-843-38	Sequence 38, Appli
6	19.2	64.0	US-08-728-956-2	Sequence 2, Appli
7	18.4	61.3	US-08-686-878A-8	Sequence 8, Appli
8	18.4	61.3	US-09-175-928-13	Sequence 13, Appli
9	18.4	61.3	US-08-936-165A-189	Sequence 189, App
10	18.4	61.3	US-09-734-674-3	Sequence 3, Appli
11	18.4	60.0	US-08-858-207A-37	Sequence 37, Appli
12	18.4	60.0	US-07-792-685D-1	Sequence 1, Appli
13	18.4	60.0	US-08-961-527-129	Sequence 129, App
14	18.4	60.0	US-09-685-853A-3	Sequence 3, Appli
15	17.8	59.3	US-09-671-317-387	Sequence 387, App
16	17.8	59.3	PCT-US93-03076-1	Sequence 1, Appli
17	17.8	59.3	US-08-916-421B-1	Sequence 1, Appli
18	17.6	58.7	US-09-368-588-1	Sequence 1, Appli
19	17.6	58.7	US-09-614-034-189	Sequence 189, App
20	17.4	58.0	US-09-675-305-5	Sequence 5, Appli
21	17.4	58.0	US-09-328-352-1822	Sequence 1822, Ap
22	17.4	58.0	US-09-107-532A-2701	Sequence 2701, Ap
23	17.4	58.0	US-09-675-305-13	Sequence 13, Appli
24	17.4	58.0	US-09-276-531-78	Sequence 78, Appli
25	17.4	58.0	US-08-868-786-5	Sequence 5, Appli
26	17.4	58.0	US-09-173-914-1	Sequence 1, Appli
27	17.4	58.0	US-09-426-290-1	Sequence 1, Appli

28	17.2	57.3	1082	1	US-08-716-301-5	Sequence 5, Appli
29	17.2	57.3	1673	3	US-09-381-849-6	Sequence 6, Appli
30	17.2	57.3	1698	4	US-09-252-991A-1191	Sequence 1191, Ap
31	17.2	57.3	1716	4	US-09-328-352-507	Sequence 507, App
32	17.2	57.3	1890	4	US-09-252-991A-1417	Sequence 1417, Ap
33	17.2	57.3	2625	1	US-08-188-228-57	Sequence 57, Appli
34	17.2	57.3	2625	1	US-08-332-643-51	Sequence 51, Appli
35	17.2	57.3	2625	1	US-08-332-638-57	Sequence 57, Appli
36	17.2	57.3	2879	4	US-08-580-031A-4	Sequence 4, Appli
37	17.2	57.3	3712	2	US-08-738-349-3	Sequence 3, Appli
38	17.2	57.3	3914	2	US-08-738-349-5	Sequence 5, Appli
39	17.2	57.3	1864	4	US-08-961-527-61	Sequence 61, Appli
40	17.2	57.3	14636	3	US-09-173-914-6	Sequence 6, Appli
41	17.2	57.3	1230025	4	US-09-188-452A-1	Sequence 1, Appli
42	17	56.7	663	3	US-08-998-416-187	Sequence 187, App
43	17	56.7	719	3	US-08-998-416-1138	Sequence 1138, Ap
44	17	56.7	856	3	US-08-998-416-289	Sequence 289, App
45	17	56.7	1041	4	US-09-453-702B-189	Sequence 189, App

# ALIGNMENTS

RESULT 1  
US-08-468-036-4/C  
Sequence 4, Application US/08468036  
Patent No. 572806  
GENERAL INFORMATION:  
APPLICANT: Demaggio, Anthony J.  
APPLICANT: Hoeckstra, Meri F.  
TITLE OF INVENTION: Materials and Methods Relating to Proteins that  
NUMBER OF INVENTIONS: Interact with Casein Kinase I  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,036  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/184,605  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 572806and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/31784  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6854 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURES:  
NAME/KEY: CDS  
LOCATION: 2050..4053  
US-08-468-036-4  
Query Match 64.7% Score 19.4, DB 1, Length 6854,



CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/184,605  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5728906and, Greeta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/31784  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3466 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-468-036-38

Query Match 64.0%; Score 19.2; DB 1; Length 3466;  
Best Local Similarity 87.5%; Pred. No. 41;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATAGAGCCGAGAAATTCAG 24  
DB 353 ATAGAGAGCCGAGAAATTCAG 330

RESULT 5  
US-08-376-843-38/c  
Sequence 38, Application US/08376843  
Patent No. 5846764  
GENERAL INFORMATION:  
APPLICANT: Demaglo, Anthony J.  
TITLE OF INVENTION: Materials and Methods Relating to Proteins  
TITLE OF INVENTION: that interact with Casein Kinase I  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/376,843  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/184,605  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5846764and, Greeta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/31784  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3466 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
US-08-376-843-38

Query Match 64.0%; Score 19.2; DB 2; Length 3466;  
Best Local Similarity 87.5%; Pred. No. 41;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATAGAGCCGAGAAATTCAG 24  
DB 353 ATAGAGAGCCGAGAAATTCAG 330

RESULT 6  
US-08-728-956-2  
Sequence 2, Application US/08728956  
Patent No. 5677175  
GENERAL INFORMATION:  
APPLICANT: Hodges, Thomas K.  
TITLE OF INVENTION: Plant Pathogen Induced Proteins  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Barnes & Thornburg  
STREET: 11 S Meridian  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,956  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Ereen, John P.  
REGISTRATION NUMBER: 38,833  
REFERENCE/DOCKET NUMBER: 3220-26119  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)231-7745  
TELEFAX: (317)231-7433  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 613 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Oryza sativa  
US-08-728-956-2

Query Match 63.3%; Score 19; DB 1; Length 613;  
Best Local Similarity 81.5%; Pred. No. 37;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 TAGAGCCGAGAAATTCAGTGCAC 30  
DB 292 TTGAGAGAGCCGAGAAATTCAGTGCAC 318

RESULT 7  
US-08-686-878A-8/c  
Sequence 8, Application US/08686878A  
Patent No. 5708157  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John

APPLICANT: Lavallie, Edward  
APPLICANT: Racie, Lisa  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Evans, Cheryl  
APPLICANT: Spaulding, Vikki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/686,878A  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 429 Base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-686-878A-8

Query Match 61.3%; Score 18.4; DB 1; Length 429;  
Best Local Similarity 78.6%; Pred. No. 62;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TATGAGGCCCAAGAAATCAGCTGAC 29  
DB 333 TACAGAGCCCAAGAAAGCAGAGAC 306

RESULT 8  
US-09-175-928-13/C  
Sequence 13, Application US/09175928A  
Patent No. 6312921  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallie, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: M., Sha  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: 6006B,AJ17A  
CURRENT APPLICATION NUMBER: US/09/175,928A  
CURRENT FILING DATE: 1998-10-20  
NUMBER OF SEQ ID NOS: 62  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 13  
LENGTH: 429  
TYPE: DNA  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: unsure  
LOCATION: (10)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (18)...(19)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (24)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (97)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (117)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (142)...(143)  
US-09-175-928-13

Query Match 61.3%; Score 18.4; DB 4; Length 429;  
Best Local Similarity 78.6%; Pred. No. 62;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TATGAGGCCCAAGAAATCAGCTGAC 29  
DB 333 TACAGAGCCCAAGAAAGCAGAGAC 306

RESULT 9  
US-08-936-165A-189  
Sequence 189, Application US/08936165A  
Patent No. 6348582  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lonetto, Michael  
APPLICANT: Nicholas, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Reichard, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
TITLE OF INVENTION: Polypeptides and their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESSES:  
ADDRESS: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimm, Edward R.  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090

TELEX:  
INFORMATION FOR SEQ ID NO: 189:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1355 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-08-936-165A-189

Query Match 61.3%; Score 18.4; DB 4; Length 1355;  
Best Local Similarity 75.9%; Pred. No. 75;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATATAGAGCCCAAGAAAATCAGCTGAC 29  
DB 972 ACACAGAGCTGACGAAAGTCAGCTGAC 1000

RESULT 10  
US-09-734-674-3/c  
Sequence 3, Application US/09734674  
Patent No. 6498022  
GENERAL INFORMATION:  
APPLICANT: WEI, Ming-Hui et al  
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: CL001018  
CURRENT APPLICATION NUMBER: US/09/734,674  
CURRENT FILING DATE: 2000-12-13  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 202001  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(202001)  
OTHER INFORMATION: n = A,T,C or G  
US-09-734-674-3

Query Match 61.3%; Score 18.4; DB 4; Length 202001;  
Best Local Similarity 78.6%; Pred. No. 1.6e+02;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATATAGAGCCCAAGAAAATCAGCTGA 28  
DB 170709 ATATAGAGCTCAATGAAACATCACTTGA 170682

RESULT 11  
US-08-858-207A-37  
Sequence 37, Application US/08858207A  
Patent No. 6348328  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Nicholas, Richard  
APPLICANT: Stodola, Robert  
TITLE OF INVENTION: No. 6348328e1 Compounds  
NUMBER OF SEQUENCES: 552  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS for Windows Version 2.0  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,207A  
FILING DATE: 09-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/017670  
FILING DATE: 14-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gilmu, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50475  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2179 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-858-207A-37

Query Match 60.0%; Score 18; DB 4; Length 2179;  
Best Local Similarity 80.8%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 ATAGAGCCCAAGAAAATCAGCTGA 28  
DB 803 ATATAGCTTATGAGATCACTGCTTA 828

RESULT 12  
US-07-792-865D-1  
Sequence 1, Application US/0792865D  
Patent No. 5646247  
GENERAL INFORMATION:  
APPLICANT: John W. Barnwell, Mary W. Galsinski,  
APPLICANT: Samuel P. Wertheimer  
TITLE OF INVENTION: MEROZOITE ANTIGENS LOCALIZED AT  
TITLE OF INVENTION: THE APICAL END OF THE PARASITE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dardy & Dardy P.C.  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
COMPUTER: IBM or IBM-compatible  
OPERATING SYSTEM: PC/MS-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/792,865D  
FILING DATE: 19911004  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US90/01849  
FILING DATE: April 3, 1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson, Arthur  
REGISTRATION NUMBER: 34,354  
REFERENCE/DOCKET NUMBER: 5966/14692-US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 527-7700  
TELEFAX: (212) 753-6237  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:



```

;
; LENGTH: 3763 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; DESCRIPTION:
; HYPOTHETICAL: No
; ANTI-SENSE: No
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: P.vivax
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL LINE:
; ORGANELL:
; IMMEDIATE SOURCE:
; LIBRARY: lambda gt 11 native P.vivax
; LIBRARY: DNA expression library
; CLONE: 5.3
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: This sequence corresponds to
; OTHER INFORMATION: Figure 1A (sheets 1-4) in the
; OTHER INFORMATION: application, as filed.
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
; US-07-792-865D-1

Query Match 60.0%; Score 18; DB 1; Length 3763;
Best Local Similarity 80.8%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATATGAGGCCCAAGAAATCAGCT 26
DB 971 ATATGAGGCCCAAGAAATCAGCT 996

RESULT 13
US-08-961-527-129/C
; Sequence 129, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:

```

```

;
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Broctec, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 129:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8512 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-129

Query Match 60.0%; Score 18; DB 4; Length 8512;
Best Local Similarity 80.8%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 ATAGAGGCCCAAGAAATCAGCTGA 28
DB 2960 ATAGAGGCCCAAGAAATCAGCTGA 2935

RESULT 14
US-09-685-853A-3/C
; Sequence 3, Application US/09685853A
; Patent No. 6479270
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: C1000871
; CURRENT APPLICATION NUMBER: US/09/685,853A
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: 60/182,194
; PRIOR FILING DATE: 2000-02-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 3
; LENGTH: 74962
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(74962)
; OTHER INFORMATION: n = A,T,C or G
; US-09-685-853A-3

Query Match 60.0%; Score 18; DB 4; Length 74962;
Best Local Similarity 80.8%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATATGAGGCCCAAGAAATCAGCT 26
DB 74050 ATATGAGGCCCAAGAAATCAGCT 74025

RESULT 15
US-09-671-317-387
; Sequence 387, Application US/09671317

```

```

; Patent No. 6528260
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIOMIMETIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62.US3.CIP
; CURRENT APPLICATION NUMBER: US/09/671,317
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 387
; LENGTH: 1000
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 497
; OTHER INFORMATION: 12-662-80 : polymorphic base G or C
; NAME/KEY: misc_binding
; LOCATION: 477..496
; OTHER INFORMATION: 12-662-80.misl, potential
; NAME/KEY: misc_binding
; LOCATION: 498..517
; OTHER INFORMATION: 12-662-80.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 418..435
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 979..1000
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 485..509
; OTHER INFORMATION: 12-662-80 potential probe
; NAME/KEY: misc_feature
; LOCATION: 24,726,739
; OTHER INFORMATION: n=a, g, c or t
; US-09-671-317-387

```

```

Query Match 59.3%; Score 17.8; DB 4; Length 1000;
Best Local Similarity 75.9%; Pred.No.1.3e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Cy 1 ATATGAGGCCAGAAAAATCAGCTGAC 29
Db 971 ATGTAGAGGACAGAAAAAAGCTTGAC 999

```

Search completed: December 18, 2003, 09:04:15  
Job time : 49.7705 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2003, 09:02:17 ; Search time 965.902 Seconds

(without alignments)  
103.495 Million cell updates/sec

Title: US-09-889-491-13

Perfect score: 30  
Sequence: 1 atatagaagcccaagaataacagctgacc 30

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapept 1.0

Searched: 221978 seqs, 1665101734 residues

Total number of hits satisfying chosen parameters: 4423956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/2/pubpna/PCN\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/2/pubpna/PCNUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq2:\*  
14: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*  
18: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22.2	74.0	2577	9 US-09-815-242-6797	Sequence 6797, Ap
2	22.2	74.0	17087	10 US-09-070-927A-47	Sequence 47, Ap
3	20.4	68.0	1923	13 US-10-027-632-255052	Sequence 255052, Ap
4	20.4	68.0	1923	14 US-10-027-632-255052	Sequence 255052, Ap
5	20	66.7	3610	10 US-09-880-107-2406	Sequence 2406, Ap
6	20	66.7	3610	13 US-09-873-367-721	Sequence 721, Ap
7	20	66.7	3702	15 US-10-084-817-30	Sequence 30, Ap
8	19.6	65.3	405	11 US-09-918-995-36137	Sequence 36137, A
9	19.6	65.3	52354	9 US-09-742-311-3	Sequence 3, Appl
10	19.4	64.7	406	9 US-09-815-343-1379	Sequence 1379, Ap
11	19.2	64.0	837	13 US-10-027-632-147079	Sequence 147079, Ap
12	19.2	64.0	837	14 US-10-027-632-147079	Sequence 147079, Ap
13	19.2	64.0	1695	13 US-10-032-585-6149	Sequence 6149, Ap
14	19.2	64.0	2169	9 US-09-815-242-6699	Sequence 6699, Ap
15	19.2	64.0	20561	10 US-09-070-927A-292	Sequence 292, Ap

C 16	19	63.3	483	9 US-09-864-761-6011	Sequence 6011, Ap
C 17	19	63.3	806	9 US-09-770-445-794	Sequence 794, Ap
C 18	19	63.3	2174	13 US-10-027-632-256360	Sequence 256360, Ap
C 19	19	63.3	2174	14 US-10-027-632-256360	Sequence 256360, Ap
C 20	18.8	62.7	133	9 US-09-770-696-601	Sequence 601, Ap
C 21	18.8	62.7	560	15 US-10-157-031-258	Sequence 258, Ap
C 22	18.8	62.7	1100	13 US-10-027-632-9964	Sequence 9964, Ap
C 23	18.8	62.7	1100	14 US-10-027-632-9964	Sequence 9964, Ap
C 24	18.8	62.7	4403	8 US-08-781-986A-78	Sequence 78, Appl
C 25	18.8	62.7	8805	10 US-09-070-927A-135	Sequence 135, Appl
C 26	18.6	62.0	504	10 US-09-479-040-6	Sequence 6, Appl
C 27	18.6	62.0	642	13 US-10-027-632-251328	Sequence 251328, Ap
C 28	18.6	62.0	642	14 US-10-027-632-251328	Sequence 251328, Ap
C 29	18.6	62.0	673	13 US-10-027-632-285692	Sequence 285692, Ap
C 30	18.6	62.0	673	14 US-10-027-632-285692	Sequence 285692, Ap
C 31	18.6	62.0	7916	10 US-09-479-040-1	Sequence 1, Appl
C 32	18.6	62.0	10046	9 US-09-764-887-458	Sequence 458, Appl
C 33	18.6	62.0	10046	15 US-10-073-961-458	Sequence 458, Appl
C 34	18.4	61.3	261	10 US-09-867-701-4483	Sequence 4483, Ap
C 35	18.4	61.3	261	10 US-09-867-701-7801	Sequence 7801, Ap
C 36	18.4	61.3	261	11 US-09-822-946-67	Sequence 67, Appl
C 37	18.4	61.3	285	10 US-09-867-701-4769	Sequence 4769, Ap
C 38	18.4	61.3	296	14 US-10-114-693-142	Sequence 142, Appl
C 39	18.4	61.3	405	10 US-09-938-842A-3609	Sequence 3609, Ap
C 40	18.4	61.3	429	14 US-10-040-916-8	Sequence 8, Appl
C 41	18.4	61.3	429	15 US-10-016-349-13	Sequence 13, Appl
C 42	18.4	61.3	441	13 US-10-027-632-182663	Sequence 182663, Ap
C 43	18.4	61.3	441	13 US-10-027-632-182664	Sequence 182664, Ap
C 44	18.4	61.3	441	14 US-10-027-632-182663	Sequence 182663, Ap
C 45	18.4	61.3	441	14 US-10-027-632-182664	Sequence 182664, Ap

#### ALIGNMENTS

RESULT 1  
US-09-815-242-6797  
Sequence 6797, Application US/09815242  
Patent No. US2002061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Cair, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6797  
LENGTH: 2577  
TYPE: DNA  
ORGANISM: Enterococcus faecalis

FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(2577)  
US-09-815-242-6797

Query Match  
Best Local Similarity 74.0%; Score 22.2; DB 9; Length 2577;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TAGAAGCCCAAGAAAATCAGTGAAC 30  
DB 1520 TAGAAGCCCAAGAAAATCAGTGAAC 1546

RESULT 2  
US-09-070-927A-47/c  
Sequence 47, Application US/09070927A  
Patent No. US20020120116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunach  
Patrick J. Dillon  
Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS Version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/070, 927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: P8369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17087 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-09-070-927A-47  
SEQUENCE DESCRIPTION: SEQ ID NO: 47:

Query Match  
Best Local Similarity 74.0%; Score 22.2; DB 10; Length 17087;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TAGAAGCCCAAGAAAATCAGTGAAC 30  
DB 11327 TAGAAGCCCAAGAAAATCAGTGAAC 11301

RESULT 3

US-10-027-632-255052/c  
Sequence 255052, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
POLYMORPHISMS IN THE HUMAN GENOME  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027, 632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1998-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 255052  
LENGTH: 1923  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-255052

Query Match  
Best Local Similarity 68.0%; Score 20.4; DB 13; Length 1923;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATATAGAGCCCAAGAAAATCAGTGAAC 30  
DB 1916 ATATAGAGCCCAAGAAAATCAGTGAAC 1887

RESULT 4  
US-10-027-632-255052/c  
Sequence 255052, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
POLYMORPHISMS IN THE HUMAN GENOME  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027, 632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1998-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 255052  
LENGTH: 1923  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-255052

Query Match 68.0%; Score 20.4; DB 14; Length 1923;  
Best Local Similarity 80.0%; Pred. No. 1.4e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATATAGAGCCCAAGAAAATCGCTGACC 30  
DB 1916 ATATAGAGCCCAAGAAAATCGATGACC 1887

## RESULT 5

US-09-880-107-2406  
Sequence 2406, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Schert, Uwe  
TITLE OF INVENTION: Gene Logic, Inc.  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2406  
LENGTH: 3610  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURES:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 M83216  
US-09-880-107-2406

Query Match 66.7%; Score 20; DB 10; Length 3610;  
Best Local Similarity 82.1%; Pred. No. 2.2e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 ATGAGAGCCCAAGAAAATCGCTGACC 30  
DB 1112 ATGAGAGCCCAAGAAAATCGCTGCC 1139

## RESULT 6

US-09-873-367C-721  
Sequence 721, Application US/09873367C  
Patent No. US2003015839A1  
GENERAL INFORMATION:  
APPLICANT: Young, Paul  
APPLICANT: Soppet, Daniel  
APPLICANT: Endress, Gregory  
APPLICANT: Augustus, Meena  
APPLICANT: Ederer, Reinhard  
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using  
FILE REFERENCE: 689290-64  
CURRENT APPLICATION NUMBER: US/09/873,367C  
CURRENT FILING DATE: 2003-04-29  
PRIOR APPLICATION NUMBER: U.S. 60/236,891  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: U.S. 60/236,842  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: U.S. 60/244,867  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: U.S. 60/245,084  
PRIOR FILING DATE: 2000-11-01  
NUMBER OF SEQ ID NOS: 1067  
SOFTWARE: PatentIn Version 3.0  
SEQ ID NO 721  
LENGTH: 3610

TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-873-367C-721

Query Match 66.7%; Score 20; DB 13; Length 3610;  
Best Local Similarity 82.1%; Pred. No. 2.2e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 ATGAGAGCCCAAGAAAATCGCTGACC 30  
DB 1112 ATGAGAGCCCAAGAAAATCGCTGCC 1139

## RESULT 7

US-10-084-817-30  
Sequence 30, Application US/10084817  
Patent No. US20030119099A1  
GENERAL INFORMATION:  
APPLICANT: Susan Stuart  
APPLICANT: Ted G. Nuchtern  
APPLICANT: Sharon R. Plon  
APPLICANT: Jason M. Shohet  
TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION  
FILE REFERENCE: PA-0046 US  
CURRENT APPLICATION NUMBER: US/10/084,817  
CURRENT FILING DATE: 2002-02-25  
PRIOR APPLICATION NUMBER: 60/270,784  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 365  
SOFTWARE: PERL Program  
SEQ ID NO 30  
LENGTH: 3702  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURES:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20030119099A1 1312325CB1  
US-10-084-817-30

Query Match 66.7%; Score 20; DB 15; Length 3702;  
Best Local Similarity 82.1%; Pred. No. 2.2e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 ATGAGAGCCCAAGAAAATCGCTGACC 30  
DB 1328 ATGAGAGCCCAAGAAAATCGCTGCC 1355

## RESULT 8

US-09-918-995-36137  
Sequence 36137, Application US/09918995  
Patent No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 36137  
LENGTH: 405  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-918-995-36137

Query Match 65.3%; Score 19.6; DB 11; Length 405;  
Best Local Similarity 84.6%; Pred. No. 2.2e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 TATAGAGCCCAAGAAAATCAGCTG 27  
DB 258 TAGAGAGCCCAAGAAAATCAGCTG 283

## RESULT 9

US-09-742-311-3  
Sequence 3, Application US/09742311  
Patent No. US20020028773A1  
GENERAL INFORMATION:  
APPLICANT: GUEGLER, Karl et al  
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
FILE REFERENCE: C000656  
CURRENT FILING DATE: 2000-12-22  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 52354  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1) .. (52354)  
OTHER INFORMATION: n = A,T,C or G  
US-09-742-311-3

Query Match 65.3%; Score 19.6; DB 9; Length 52354;  
Best Local Similarity 84.6%; Pred. No. 5.2e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 ATGAGAGCCCAAGAAAATCAGCTGA 28  
DB 30034 ACAGATGCTCAATATAAATCAGCTGA 30059

## RESULT 10

US-09-815-343-1379/c  
Sequence 1379, Application US/09815343  
Patent No. US20010055596A1  
GENERAL INFORMATION:  
APPLICANT: Meagher, Madeleine  
APPLICANT: Xu, Jiangchun  
APPLICANT: King, Gordon E.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER  
FILE REFERENCE: 210121.504  
CURRENT APPLICATION NUMBER: US/09/815,343  
CURRENT FILING DATE: 2001-03-22  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1379  
LENGTH: 406  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1) .. (406)  
OTHER INFORMATION: n = A,T,C or G  
US-09-815-343-1379

Query Match 64.7%; Score 19.4; DB 9; Length 406;  
Best Local Similarity 79.3%; Pred. No. 2.7e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 AATTAAGAGCCCAAGAAAATCAGCTGAC 29  
DB 154 AATTAAGAGTACCAAGAAAATCAGCTGAC 126

## RESULT 11

US-10-027-632-147079/c  
Sequence 147079, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
TITLE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 147079  
LENGTH: 837  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-147079

Query Match 64.0%; Score 19.2; DB 13; Length 837;  
Best Local Similarity 80.8%; Pred. No. 3.6e+02;  
Matches 21; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 5 AGAAGCCCAAGAAAATCAGCTGACC 30  
DB 80 AGAATCAAGAAAATCAGCTGACC 55

## RESULT 12

US-10-027-632-147079/c  
Sequence 147079, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
TITLE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 147079  
LENGTH: 837  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-147079

RESULT 14  
US-09-815-242-6699  
Sequence 6699, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA-011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625

```

1 STRANDEDNESS: double
2
3 TOPOLOGY: linear
4
5 SEQUENCE DESCRIPTION: SEQ ID NO: 292:

```

Query Match 64.0%; Score 19.2; DB 10; Length 20561;  
 Best Local Similarity 87.5%; Pred. No. 6.4e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 6 GAAAGCCCAAGAAAAATCGCTGAC 29  
 |||||  
 Db 19006 GAAAGCCCAAGAAAAATCGCTGAC 18983

Search completed: December 18, 2003, 10:22:57  
 Job time : 968.902 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2003, 09:02:17, Search time 1912.62 Seconds

(without alignments)  
381.222 Million cell updates/sec

Title: US-09-889-491-13

Perfect score: 30

Sequence: 1 atatagaagcccaagaataacagctgacc 30

Scoring table:

IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estba:  
2: em\_esthum:  
3: em\_estin:  
4: em\_estmu:  
5: em\_estcov:  
6: em\_estpl:  
7: em\_estro:  
8: em\_hlc:  
9: gb\_est1:  
10: gb\_est2:  
11: gb\_hlc:  
12: gb\_est3:  
13: gb\_est4:  
14: gb\_est5:  
15: em\_estfun:  
16: em\_estom:  
17: em\_gss\_hum:  
18: em\_gss\_inv:  
19: em\_gss\_pln:  
20: em\_gss\_vit:  
21: em\_gss\_fun:  
22: em\_gss\_mam:  
23: em\_gss\_mus:  
24: em\_gss\_pro:  
25: em\_gss\_rtd:  
26: em\_gss\_phg:  
27: em\_gss\_vil:  
28: gb\_gss1:  
29: gb\_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	22	73.3	1086 10	BE035293 NM050505 M
2	21.6	72.0	454 12	BM488025 9520322B05
3	21.6	72.0	477 10	BG266188 1000086G0
4	21.6	72.0	552 9	AT714933 606024G06

5	21.6	72.0	559 13	BU049916	BU049916 111017B0
6	21.6	72.0	645 13	BU044878	BU044878 PP IEa002
7	21.6	72.0	649 11	AY112554	AY112554 Zea may
8	21.4	71.3	940 29	CC304082	CC304082 CH261-77A
9	21.2	70.7	664 28	AO383884	AO383884 RPT11-12
10	21.2	70.0	411 13	BU948624	BU948624 int0a11.x
11	21.1	70.0	632 10	BU048787	BU048787 BB048787
12	21.1	70.0	700 28	BH833330	BH833330 BACPP4-B2
13	20.6	68.7	344 28	BH761049	BH761049 BMBAC318G
14	20.6	68.7	730 28	B2428389	B2428389 BONEW78TF
15	20.4	68.0	498 28	AO254234	AO254234 CP00718A
16	20.4	68.0	570 12	B1321439	B1321439 s8f11a05
17	20.4	68.0	778 29	CC081774	CC081774 CSU-K33F
18	20.4	68.0	1317 12	B1490262	B1490262 603032074
19	20.2	67.3	1229 29	CC030312	CC030312 CH261-59K
20	20.2	66.7	209 10	BB324020	BB324020 BB324020
21	20.2	66.7	239 9	AV236858	AV236858 AV236858
22	20.2	66.7	281 10	BF476902	BF476902 naa55f08
23	20.2	66.7	282 10	BF476962	BF476962 naa56h09
24	20.2	66.7	286 14	F07917	F07917 HSC2LE101 n
25	20.2	66.7	288 10	BF063480	BF063480 7124h11.x
26	20.2	66.7	301 10	BF105699	BF105699 BB105699
27	20.2	66.7	301 10	BF063453	BF063453 7124c06.x
28	20.2	66.7	309 10	BF063598	BF063598 7123g11.x
29	20.2	66.7	322 10	BF060828	BF060828 7127f09.x
30	20.2	66.7	348 10	BF000738	BF000738 7143e07.x
31	20.2	66.7	365 10	BF001035	BF001035 7145g07.x
32	20.2	66.7	371 14	R12015	R12015 Y651906.x1
33	20.2	66.7	381 14	N73266	N73266 yw54h10.x1
34	20.2	66.7	387 28	AZ831913	AZ831913 2M0111149
35	20.2	66.7	388 10	BF477200	BF477200 naa64b09
36	20.2	66.7	415 10	BF060868	BF060868 7128g06.x
37	20.2	66.7	415 14	R19486	R19486 y921d06.x1
38	20.2	66.7	416 10	BE394869	BE394869 6013106626
39	20.2	66.7	428 10	BF060807	BF060807 7127a05.x
40	20.2	66.7	444 2	HSW071130	HSW071130 Homo sapi
41	20.2	66.7	458 9	BF060813	BF060813 7127b04.x
42	20.2	66.7	461 9	AM083647	AM083647 xc49a01.x
43	20.2	66.7	461 14	HS3171	HS3171 yq83e01.x1
44	20.2	66.7	465 10	BE150759	BE150759 na102d05
45	20.2	66.7	476 10	BE588622	BE588622 7110g07.x

#### ALIGNMENTS

RESULT 1  
BE035293/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BE035293 1086 bp mRNA linear EST 07-JUN-2000  
NM050505 NM Mesembryanthemum crystallinum cDNA 5', mRNA sequence.  
BE035293  
BE035293.1 GI:8330205  
EST.  
Mesembryanthemum crystallinum (common iceplant)  
Mesembryanthemum crystallinum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Caryophyllales; Rosales; Rosaceae; Mesembryanthemum.  
1 (bases 1 to 1086)  
Bohner, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M., Ferreira, H., Kawasaki, S., McColough, A., Michalowski, C.B., Palacios, C., Scara, G., Wheeler, M. and Zepeda, G.R.  
Functional Genomics of Plant Stress Tolerance  
Unpublished  
Contact: Michalowski, C.B.  
University of Arizona  
Bio Sciences West room 513, Tucson, AZ 85721, USA  
Tel: 520-621-7982  
Fax: 520-621-1697  
Email: chmou.arizona.edu  
Best Blastx match: 'gi|1946367 (U93215) unknown protein [Arabidopsis thaliana] 249 2e-65'. An open reading frame exists.  
Insert Length: 1 Std Error: 0.00.

## FEATURES

## source

Location/Qualifiers

1.1086

/organism="Mesembryanthemum crystallinum"

/mol\_type="mRNA"

/db\_xref="taxon:3544"

/cell\_type="epidermal bladder cells"

/dev\_stage="12 weeks old"

/clone\_lib="WM"

/note="Vector: Bluescript SK+, Site 1: EcoRI, Site 2: XhoI"

; Plants stressed 6 weeks in 500mM NaCl"

/lab\_host="DH10B"

/clone\_lib="952 - BMS tissue from Walbot Lab (reduced rRNA"

)

/note="Vector: pUC19, Site 1: EcoRI, Site 2: EcoRI, The"

library was prepared by George Rudenko using poly (A)

selected RNA and Universal Riboclone cDNA Synthesis System

(Promega). cDNA was synthesized using both random and

oligo(dT) primers in separate reactions and equipped with

EcoRI adaptors. Library was size-fractionated on agarose

gels (for insert size &gt;400bp) and non-directionally cloned

into EcoRI-digested pUC19 vector. Blue/white selection on

carbenicillin-containing plates was used to recover

positive clones."

119 a 130 c 99 g 106 t

BASE COUNT

ORIGIN

119 a 130 c 99 g 106 t

BASE COUNT

ORIGIN

119 a 130 c 99 g 106 t

BASE COUNT

ORIGIN

119 a 130 c 99 g 106 t

BASE COUNT

ORIGIN

119 a 130 c 99 g 106 t

BASE COUNT

ORIGIN

119 a 130 c 99 g 106 t

BASE COUNT

ORIGIN

119 a 130 c 99 g 106 t

BASE COUNT

ORIGIN

119 a 130 c 99 g 106 t

BASE COUNT

## Query Match

## Best local Similarity

Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 161

2 TATAGAGCCCAAGAGAGCTGAC 29

161 TATAGAGCCCAAGAGAGCTGAC 188

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLES

AUTHORS

COMMENT

JOURNAL

CONTACT: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 723 2227

Email: walbot@stanford.edu

Plate: 1000086 row: G column: 04.

Location/Qualifiers

1.477

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="dbEST:606024G06.x2"

/db\_xref="taxon:4577"

/clone\_lib="1000 - Unigene I from Maize Genome Project"

/note="This library represents the unique ESTs found in

the first round of EST sequencing at Stanford University

for the maize genome project. Sequences are present from

libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683

, 687, 707, and 945. Contigs were assembled using TIGR's

CAP program and a representative EST from each contig was

selected for the Unigene set. All singlets were also

selected."

116 a 143 c 113 g 105 t

BASE COUNT

ORIGIN

72.0%; Score 21.6; DB 12; Length 454;

85.7%; Pred. No. 8.6e+02;

0; Mismatches 4; Indels 0; Gaps 0;

Db 161

2 TATAGAGCCCAAGAGAGCTGAC 29

161 TATAGAGCCCAAGAGAGCTGAC 188

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLES

AUTHORS

COMMENT

JOURNAL

CONTACT: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 723 2227

Email: walbot@stanford.edu

Plate: 1000086 row: G column: 04.

Location/Qualifiers

1.477

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="dbEST:606024G06.x2"

/db\_xref="taxon:4577"

/clone\_lib="1000 - Unigene I from Maize Genome Project"

/note="This library represents the unique ESTs found in

the first round of EST sequencing at Stanford University

for the maize genome project. Sequences are present from

libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683

, 687, 707, and 945. Contigs were assembled using TIGR's

CAP program and a representative EST from each contig was

selected for the Unigene set. All singlets were also

selected."

116 a 143 c 113 g 105 t

BASE COUNT

ORIGIN

72.0%; Score 21.6; DB 12; Length 454;

85.7%; Pred. No. 8.6e+02;

0; Mismatches 4; Indels 0; Gaps 0;

Db 161

2 TATAGAGCCCAAGAGAGCTGAC 29

161 TATAGAGCCCAAGAGAGCTGAC 188

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLES

AUTHORS

COMMENT

JOURNAL

CONTACT: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 723 2227

Email: walbot@stanford.edu

Plate: 1000086 row: G column: 04.

Location/Qualifiers

1.477

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="dbEST:606024G06.x2"

/db\_xref="taxon:4577"

/clone\_lib="1000 - Unigene I from Maize Genome Project"

/note="This library represents the unique ESTs found in

the first round of EST sequencing at Stanford University

for the maize genome project. Sequences are present from

libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683

, 687, 707, and 945. Contigs were assembled using TIGR's

CAP program and a representative EST from each contig was

selected for the Unigene set. All singlets were also

selected."

116 a 143 c 113 g 105 t

BASE COUNT

ORIGIN

72.0%; Score 21.6; DB 12; Length 454;

85.7%; Pred. No. 8.6e+02;

0; Mismatches 4; Indels 0; Gaps 0;

Db 161

2 TATAGAGCCCAAGAGAGCTGAC 29

161 TATAGAGCCCAAGAGAGCTGAC 188

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLES

AUTHORS

COMMENT

JOURNAL

CONTACT: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 723 2227

Email: walbot@stanford.edu

Plate: 1000086 row: G column: 04.

Location/Qualifiers

1.477

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="dbEST:606024G06.x2"

/db\_xref="taxon:4577"

/clone\_lib="1000 - Unigene I from Maize Genome Project"

/note="This library represents the unique ESTs found in

the first round of EST sequencing at Stanford University

for the maize genome project. Sequences are present from

libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683

, 687, 707, and 945. Contigs were assembled using TIGR's

CAP program and a representative EST from each contig was

selected for the Unigene set. All singlets were also

selected."

116 a 143 c 113 g 105 t

BASE COUNT

ORIGIN

## REFERENCE

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoidae; Andropogoneae; Zea.

## AUTHORS

Walbot, V.

## TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford

## JOURNAL

Unpublished

## COMMENT

Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 606024 row: G column: 06.  
Location/Qualifiers

## FEATURES

## SOURCE

1. .552  
/organism="Zea mays"  
/mol\_type="mRNA"  
/culivar="Zm01043"  
/db\_xref="taxon:4577"  
/tissue\_type="mixed"  
/dev\_stage="ear length from 0.5 cm - 2.0 cm"  
/lab\_host="XLOLR (Stratagene)"  
/clone\_lib="606 - Ear tissue cDNA library from Schmidt  
lab"  
/note="Organ: immature ear; Vector: pBK-CMV; Site: 1; EcoRI  
; Site\_2: XhoI; Mixed ear tissue cDNA library from Schmidt  
lab"  
BASE COUNT 137 a 149 c 126 g 139 t 1 others  
ORIGIN

## Query Match

Best Local Similarity 72.0%; Score 21.6; DB 9; Length 552;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

## QY

2 TATAGAGCCCAAGAAATCAGCTGAC 29

## Db

182 TATAGAGCCCAAGAGACTCAGCTGAC 209

RESULT 5 559 bp mRNA linear EST 26-AUG-2002  
LOCUS BU049916 1111017B02.y1 1111 - Unigene III from Maize Genome Project Zea mays  
DEFINITION cDNA, mRNA sequence.

## ACCESSION

BU049916  
BU049916.1 GI:22489993

## VERSION

EST.

## KEYWORDS

Zea mays  
Zea mays

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoidae; Andropogoneae; Zea.

## REFERENCE

1 (bases 1 to 559)

## AUTHORS

Walbot, V.

## TITLE

Maize ESTs from various cDNA libraries sequenced at Stanford

## JOURNAL

Unpublished

## COMMENT

Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 111017 row: B column: 02.  
Location/Qualifiers

## FEATURES

## SOURCE

1. .559  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="dbEST:952032505.y1"  
/db\_xref="taxon:4577"

/clone\_lib="1111 - Unigene III from Maize Genome Project"  
/note="This library represents the unique genes found in  
the third round of EST sequencing at Stanford University  
for the maize genome project. Sequences are present from  
library 952. Contigs were assembled using ZmBAssembler  
and 2 representatives from each contig were selected for  
the unigene set. All singlets were also selected."

## BASE COUNT

149 a 159 c 127 g 124 t

## Query Match

Best Local Similarity 72.0%; Score 21.6; DB 13; Length 559;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

## QY

2 TATAGAGCCCAAGAAATCAGCTGAC 29

## Db

152 TATAGAGCCCAAGAGACTCAGCTGAC 179

RESULT 6 645 bp mRNA linear EST 26-AUG-2002  
LOCUS BU044878 PP IEA0020M10F Peach developing fruit mesocarp Prunus persica cDNA  
DEFINITION clone PP IEA0020M10F, mRNA sequence.

## ACCESSION

BU044878  
BU044878.1 GI:22484955

## VERSION

EST.

## KEYWORDS

Prunus persica (peach)

## SOURCE

Prunus persica

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosid 1; Rosales; Rosaceae; Amygdaloideae; Prunus.  
1 (bases 1 to 645)  
Callahan, A., Palmer, M., Main, D., Wing, R. and Abbott, A.  
Peach Model Genome for Rosaceae  
Unpublished  
Contact: Abbott, A.  
Dept of Genetics and Biochemistry  
Clemson University  
122 Long Hall, Clemson University, Clemson, SC 29634, USA  
Tel: 864 656 3060  
Fax: 864 656 6879  
Email: aalbert@clemson.edu  
Total High Quality bases = 603  
Seq primer: TATAGAGCTCAGCTGACG  
High quality sequence stop: 645.  
Location/Qualifiers

## QY

3 ATGAGAGCCCAAGAAATCAGCTGAC 30

## Db

152 TATAGAGCCCAAGAGACTCAGCTGAC 179

RESULT 6 645 bp mRNA linear EST 26-AUG-2002  
LOCUS BU044878 PP IEA0020M10F Peach developing fruit mesocarp Prunus persica cDNA  
DEFINITION clone PP IEA0020M10F, mRNA sequence.

## ACCESSION

BU044878  
BU044878.1 GI:22484955

## VERSION

EST.

## KEYWORDS

Prunus persica (peach)

## SOURCE

Prunus persica

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosid 1; Rosales; Rosaceae; Amygdaloideae; Prunus.  
1 (bases 1 to 645)  
Callahan, A., Palmer, M., Main, D., Wing, R. and Abbott, A.  
Peach Model Genome for Rosaceae  
Unpublished  
Contact: Abbott, A.  
Dept of Genetics and Biochemistry  
Clemson University  
122 Long Hall, Clemson University, Clemson, SC 29634, USA  
Tel: 864 656 3060  
Fax: 864 656 6879  
Email: aalbert@clemson.edu  
Total High Quality bases = 603  
Seq primer: TATAGAGCTCAGCTGACG  
High quality sequence stop: 645.  
Location/Qualifiers

## QY

3 ATGAGAGCCCAAGAAATCAGCTGAC 30

## Db

152 TATAGAGCCCAAGAGACTCAGCTGAC 179

RESULT 6 645 bp mRNA linear EST 26-AUG-2002  
LOCUS BU044878 PP IEA0020M10F Peach developing fruit mesocarp Prunus persica cDNA  
DEFINITION clone PP IEA0020M10F, mRNA sequence.

## ACCESSION

BU044878  
BU044878.1 GI:22484955

## VERSION

EST.

## KEYWORDS

Prunus persica (peach)

## SOURCE

Prunus persica

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosid 1; Rosales; Rosaceae; Amygdaloideae; Prunus.  
1 (bases 1 to 645)  
Callahan, A., Palmer, M., Main, D., Wing, R. and Abbott, A.  
Peach Model Genome for Rosaceae  
Unpublished  
Contact: Abbott, A.  
Dept of Genetics and Biochemistry  
Clemson University  
122 Long Hall, Clemson University, Clemson, SC 29634, USA  
Tel: 864 656 3060  
Fax: 864 656 6879  
Email: aalbert@clemson.edu  
Total High Quality bases = 603  
Seq primer: TATAGAGCTCAGCTGACG  
High quality sequence stop: 645.  
Location/Qualifiers

Db 135 ATTGAAGCCCAAGAAAATGACCTGACC 108

RESULT 7  
LOCUS AY112254  
DEFINITION Zea mays CL9004\_1 mRNA sequence.  
ACCESSION AY112254  
VERSION AY112254.1 GI:21216844  
KEYWORDS HTC.  
SOURCE Zea mays  
ORGANISM Zea mays

REFERENCE  
AUTHORS Hailey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whiteitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.  
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes  
JOURNAL Unpublished (2002)  
REFERENCE 2 (bases 1 to 649)  
AUTHORS Coe, E.H.  
TITLE Direct Submission  
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA  
COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, [maizemap.org](http://maizemap.org); ZmDB, [www.zmdb.iastate.edu](http://www.zmdb.iastate.edu); TIGR, [www.tigr.org](http://www.tigr.org); or NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: [www.zmdb.iastate.edu](http://www.zmdb.iastate.edu).

FEATURES  
source  
location/Qualifiers  
1..649  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="MaizeDB:633041"  
/db\_xref="taxon:4577"  
/clone\_lib="Maize Mapping Project/DuPont Consensus Library"  
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 175 a 146 c 131 g 172 t 25 others

ORIGIN

Query Match 72.0%; Score 21.6; DB 11; Length 649;  
Best Local Similarity 85.7%; Pred. No. 9.2e+02;  
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 ATGAAGCCCAAGAAAATGACCTGACC 30  
|||||  
Db 615 ATATAGCCCAAGATTAATGACCTGACC 588  
|||||

RESULT 8  
LOCUS CC304082  
DEFINITION CH261-77A22.RM1.1 CH261 Gallus gallus genomic clone CH261-77A22, genomic survey sequence.  
ACCESSION CC304082  
VERSION CC304082.1 GI:30675523  
KEYWORDS GSS.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE  
1 (bases 1 to 940)  
Location/Qualifiers

AUTHORS Krenitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.  
TITLE Gallus gallus BAC End Reads  
JOURNAL Unpublished  
COMMENT Contact: Richard K. Wilson  
Genome Sequencing Center  
Washington University School of Medicine  
Email: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
Insert Length: 182000 Std Error: 0.00  
Seq primer: RM1 TAGACTCACTATAGGAGAG  
Class: BAC ends  
High quality sequence start: 35  
High quality sequence stop: 681.  
Location/Qualifiers  
1..940  
/organism="Gallus gallus"  
/mol\_type="genomic DNA"  
/strain="Red Jungle Fowl"  
/db\_xref="taxon:9031"  
/clone="CH261-77A22"  
/sex="Female"  
/cell\_line="UCDP001, inbred 256"  
/note="Vector: pFARBAC2.1; Site 1: EcoRI; Site 2: EcoRI; CH261 Female Chicken library - For library and clone ordering information: <http://www.choil.org/bacpac>"

BASE COUNT 304 a 192 c 129 g 313 t 2 others

ORIGIN

Query Match 71.3%; Score 21.4; DB 29; Length 940;  
Best Local Similarity 95.7%; Pred. No. 1.2e+03;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AGAAGCCCAAGAAAATGACCTG 27  
|||||  
Db 57 AGAAGCCCAAGAAAATGACCTG 35  
|||||

RESULT 9  
LOCUS AQ383884  
DEFINITION RPCI11-122J17.TV RPCI-11 Homo sapiens genomic clone RPCI-11-122J17, genomic survey sequence.  
ACCESSION AQ383884  
VERSION AQ383884.1 GI:4354907  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 664)  
Zhaio, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
Unpublished  
Other\_GSSs: RPCI11-122J17.TV  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Bukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [hbeet@igf.org](mailto:hbeet@igf.org)  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong ([pieterdejong@med.buffalo.edu](mailto:pieterdejong@med.buffalo.edu)). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: 8f6  
Class: BAC ends  
Location/Qualifiers

FEATURES

SOURCE

1..664  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:754696"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-123J17"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_id="RPC1-11"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC11 Human Male BAC Library"

BASE COUNT 274 a 113 c 111 g 166 t

ORIGIN

Query Match 70.7%; Score 21.2; DB 28; Length 664;  
Best Local Similarity 88.5%; Fred. No. 1.3e+03;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATAGAGCCCAAGAAATCAGCTGA 28  
|||||  
66 ATGAGAGCCCAAGAAATCAGCTCA 121

Db

RESULT 10 411 bp mRNA linear EST 21-OCT-2002  
LOCUS BU948624  
DEFINITION i070a11.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6127389 3',  
mRNA sequence.  
ACCESSION BU948624  
VERSION BU948624.1 GI:24199975  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 411)  
Melton,D., Brown,J., Kenty,G., Permult,A., Lee,C., Kaestner,K.,  
Lemisha,I., Scaerza,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Blisstein,A.,  
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,O., Cardenas  
,M., Gibbons,M., McCann,R., Cole,R., Tsagarisshwill,R., Williams,T.,  
Jackson,Y. and Bowers,Y.  
Endocrine Pancreas Consortium  
Unpublished  
TITLE  
JOURNAL  
COMMENT  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@hjh.harvard.edu  
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center for information on  
obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 385.  
Location/Qualifiers  
1..411  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6127389"  
/tissue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"  
/clone\_id="HR85 islet"  
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1:  
NotI; Site\_2: XhoI; cDNA made by oligo-dT priming  
Size selected on agarose gel. Average insert size ~1kb. 5'  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permutt Lab), Washington University

SOURCE

2 TATAGAGCCCAAGAAATCAGCTGACC 30  
|||||  
369 TATATATCTTAGGAAATCAGCTGACC 397

Db

RESULT 11 692 bp mRNA linear EST 31-AUG-2001  
LOCUS BB048787/c  
DEFINITION BB048787 RIKEN full-length enriched, adult male olfactory bulb Mus  
musculus cDNA clone 6430596G11 3', mRNA sequence.  
ACCESSION BB048787  
VERSION BB048787.2 GI:15404011  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 692)  
Arikawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Komno,H., Kouda  
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki  
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arikawa,T., et al. 2001)  
Unpublished  
TITLE  
JOURNAL  
COMMENT  
On Jun 11, 2000 this sequence version replaced gi:8455935.  
Contact: Yoshinori Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-research.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura  
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Komno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa  
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arikawa,T.,  
Ishii,Y. and Hayashizaki,Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.  
Funct. Genomics 2 pre, 172-186 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken

School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110, E-mail: hinoue@wgate.wustl.edu, Tel:  
314-362-1916, Fax: 314-747-2692."



Search completed: December 16, 2003, 11:35:01  
Job time : 1920.62 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: December 18, 2003, 09:02:17 ; Search time 1343.11 Seconds

(without alignments)  
913.763 Million cell updates/sec

Title: US-09-889-491-13

Perfect score: 30

Sequence: 1 attatagaagcccaagaataatcagctgacc 30

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 288871 segs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank1.\*

1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: gb\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rnd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vit:\*  
38: em\_sy:\*  
39: em\_hggo\_hum:\*  
40: em\_hggo\_mus:\*  
41: em\_hggo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	30	6	AX028510 Sequence
2	30	100.0	2415	9	HUMBNSEP01
3	23.8	79.3	222706	2	AC110652 Rattus no
4	23.8	79.3	235996	2	AC096147 Rattus no
5	23.8	79.3	266204	2	AC131002 Rattus no
6	23.2	77.3	86491	9	AC004810 Homo sapi
7	23.2	77.3	195296	6	AC120085 Rattus no
8	22.2	74.0	96595	6	AX65617 Sequence
9	22.2	74.0	139873	9	AC010612 Homo sapi
10	22.2	74.0	153954	9	AL354862 Human DNA
11	22.2	74.0	169059	9	AC009514 Homo sapi
12	22.2	74.0	169060	9	AC087856 Homo sapi
13	22.2	74.0	181607	9	AC021581 Homo sapi
14	22.2	74.0	304454	1	AB016956 Enterococ
15	21.6	72.0	87283	2	AC022856 Homo sapi
16	21.6	72.0	122181	9	AF222686 Homo sapi
17	21.6	72.0	171901	2	AC027240 Homo sapi
18	21.6	72.0	176364	2	AC126322 Homo sapi
19	21.4	71.3	235649	2	AC094592 Rattus no
20	21.4	71.3	253431	2	AC094604 Rattus no
21	21.4	71.3	327843	2	AC114851 Rattus no
22	21.2	70.7	65842	2	AC103866 Homo sapi
23	21.2	70.7	182527	9	AL590115 Human DNA
24	21.2	70.0	67991	9	AC006363 Homo sapi
25	21.1	70.0	122680	9	AL133510 Human DNA
26	21.1	70.0	163866	5	AL929590 Zebrafish
27	21.1	70.0	180236	2	BX248103 Dario rer
28	21.1	70.0	206773	2	AC010932 Homo sapi
29	21.1	70.0	222874	2	AC105481 Rattus no
30	20.8	69.3	151900	9	AC096644 Homo sapi
31	20.8	69.3	167346	9	AC096640 Homo sapi
32	20.6	68.7	106526	9	AC104772 Homo sapi
33	20.6	68.7	126349	2	AC101076 Mus muscu
34	20.6	68.7	167108	9	AL161666 Human chr
35	20.6	68.7	169272	2	AC016843 Homo sapi
36	20.6	68.7	187810	2	AC117939 Mus muscu
37	20.6	68.7	243659	2	AC128597 Rattus no
38	20.6	68.7	243604	2	AC136662 Rattus no
39	20.6	68.7	258719	2	AC106349 Rattus no
40	20.4	68.0	271676	2	AC097832 Rattus no
41	20.4	68.0	2000	6	AX654979 Sequence
42	20.4	68.0	6816	5	AX251443 Sequence
43	20.4	68.0	6881	5	CHRCYSTAT M95725 Gallus dome
44	20.4	68.0	16460	3	CET1358 Z81121 Caenorhabdi
45	20.4	68.0	115413	2	AC135427 Oryza sat

## ALIGNMENTS

RESULT 1  
AX028510 LOCUS 30 bp DNA linear PAT 16-SBP-2000  
DEFINITION Sequence 13 from Patent WO0042216.  
AX028510 ACCESSION  
VERSION AX028510.1 GI:10189693  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Kusk, P.  
TITLE Genetic predisposition  
JOURNAL Patent: WO 0042216-A 13 20-JUN-2000;



OSTROMETER BIOTECH AS (DK) ; KISK PHILIP (DK)

FEATURES  
source  
Location/Qualifiers  
1.30  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT  
14 a 7 c 5 g 4 t

Query Match  
Best Local Similarity 100.0%; Score 30; DB 6; Length 30;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  
1 ATATTGAAAGCCCAAGAAAATCAGCTGACC 30  
1 ATATTGAAAGCCCAAGAAAATCAGCTGACC 30

Db  
1 ATATTGAAAGCCCAAGAAAATCAGCTGACC 30

RESULT 2  
LOCUS HUMENSP01 2415 bp DNA linear PRI 29-SEP-1994  
DEFINITION Human bone sialoprotein (BNSP) gene, exon 1.  
ACCESSION L24756.1 GI:438617  
VERSION L24756.1 GI:438617  
KEYWORDS sialoprotein.  
SEGMENT 1 of 4  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Kim,R.H., Shapiro,H.S., Li,J.J., Wranz,J.L. and Sodek,J.  
TITLE Characterization of the human bone sialoprotein (BSP) gene and its promoter sequence  
JOURNAL Matrix Biol. 14 (1), 31-40 (1994)  
MEDLINE 94340202  
PMID 8061918  
COMMENT  
FEATURES  
source  
Location/Qualifiers  
1.2415  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/map="Unassigned"  
2179..2260  
/gene="BNSP"  
/number=1  
exon  
BASE COUNT 890 a 437 c 423 g 665 t

Query Match  
Best Local Similarity 100.0%; Score 30; DB 9; Length 2415;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  
1 ATATTGAAAGCCCAAGAAAATCAGCTGACC 30  
1 ATATTGAAAGCCCAAGAAAATCAGCTGACC 1510

Db  
1 ATATTGAAAGCCCAAGAAAATCAGCTGACC 1510

RESULT 3  
LOCUS AC110652 222706 bp DNA linear HTG 19-NOV-2002  
DEFINITION Rattus norvegicus clone CH230-83K21, WORKING DRAFT SEQUENCE, 4  
unordered pieces.  
ACCESSION AC110652  
VERSION AC110652.5 GI:25100613  
KEYWORDS HTG; HTGS; PHASE1; HTGS; DRAFT; HTGS; FULLTOP.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1 (bases 1 to 222706)

AUTHORS  
Mazny,D,Marle,Melzker,M,See,Abramzon,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alstbrooks,S., Amla,A., Arguano,D.,  
Aryalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
Baldwin,D., Bandaranatne,D., Barber,M., Barnstead,M., Benhmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burck,K., Cavazos,I., Ceasar,H., Chen,A.,  
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chen,A.,  
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Dlyra,K.,  
Diaper,H., Dugan-Rochin,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,P.,  
Fraser,C.M., Gabisti,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,  
Guarinate,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,  
Harvey,X., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,  
Hollins,B., Howells,S., Hu,Y.S., Hume,J., Idlebird,D., Jackson,A.,  
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,T.,  
Liu,J., Liu,M., Liu,Y., London,P., Longacre,S., Lopez,J.,  
Lorenson,W., Louiege,H., Lozano,R.J., Lu,X., Ma,V.,  
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenan,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
Nwakoeleneh,O., Okunodu,G., Oltununguon,A., Pal,S., Parks,K.,  
Pasternek,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
Plopper,F., Polidexter,A., Popovic,D., Pritts,E., Pu,L.,  
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.O.,  
Sanders,W., Saverly,G., Scherer,S., Scott,G., Shattman,S., Shen,H.,  
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajz,D.,  
Shed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soza,J.,  
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tinney,A., Tjofas,Z., Umami,K.,  
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,  
Williams,G., Willson,R., Wlezyk,R., Woodson,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,D., Yoon,V.,  
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., Von  
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Unpublished  
2 (bases 1 to 222706)  
Worley,K.C.  
Submitted (14-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 222706)  
Rat Genome Sequencing Consortium.  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23101016.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.bsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

# Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project information

Center project name: GSPZ

Center clone name: CH230-83K21

Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 21484 bases at least Q40

Consensus quality: 216820 bases at least Q30

Consensus quality: 218105 bases at least Q20

Estimated insert size: 221589; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently

consists of 4 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 17901: contig of 17901 bp in length

17902 18001: gap of unknown length

18002 217149: contig of 199148 bp in length

217150 217249: gap of unknown length

217250 221403: contig of 4154 bp in length

221404 221503: gap of unknown length

221504 222706: contig of 1203 bp in length.

Location/Qualifiers

1. 222706

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-83K21"

/complement(195782..196630)

/note="clone boundary"

clone\_end:Sp5

site: end:sequence:BH274348"

end:sequence:BH274348"

BASE COUNT 65480 a 44575 c 46422 g 62789 t 3440 others

ORIGIN

Query Match 79.3% Score 23.8; DB 2; Length 222706;

Best Local Similarity 92.6%; Pred. No. 19;

Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATGAGAGCCCAAGAAAATCAGCTG 27

Db 104561 ACATGAGAGCCCAAGAAAATCAGCTG 104587

RESULT 4

AC096147 235996 bp DNA linear HTG 10-MAY-2003

LOCUS Rattus norvegicus clone CH230-10H14, WORKING DRAFT SEQUENCE.

AC096147

AC096147.6 GI:30522612

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus

(bases 1 to 235996)

Murphy, D. Marie., Metzger, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsprouck, S., Amth, A., Anguiano, D.,

Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, N., Barnstead, F.,

Biswal, K., Blair, J., Blankenburg, K., Bluth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Buttrill, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Center, A.,

Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Devilla, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Davila, O., Denison, S., Deramo, C., Ding, Y., Din, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregiorgis, E., Geer, K., Gill, R., Giddy, M., Gier, W., Givara, M.,

Gunaratne, P., Haaland, N., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, J., Kovar, C.,

Kovacs, C., Kraft, C.L., Labow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorenz, H., Louised, H., Lozano, R.J., Lu, X., Ma, J.,

Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,

Milosevic, A., Milner, G., Ming, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munizaga, M., Murphy, M., Nair, L.,

Nankarav, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

MacKellmeh, O., Okunou, G., Olamposgon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,

Plopper, F., Polidexter, A., Popovic, D., Plimms, E., Pu, L.,

Razo, M., Quiroz, J., Rachlin, B., Reeves, K., Register, M.A., Reich, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Rutz, S.J.,

Sanders, M., Saverly, G., Scherer, S., Scott, G., Shattman, S., Shen, H.,

Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smas, D.,

Shed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J.,

Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usman, K.,

Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,

Wang, Q., Wang, S., Warren, J., Warren, J., Wei, X., White, P.,

Williams, G., Willson, R., Wleczky, R., Wooden, H., Wozley, K.,

Wright, D., Wright, R., Yu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von

Weinert, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,

Weinstock, G., and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 235996)

Worley, K.C.

Direct Submission

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 235996)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24818200.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

Genome Center  
Center: Baylor College of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
Project information  
Center project name: GENN  
Center clone name: CH230-10H14

Assembly Statistics  
Consensus quality: 224560 bases at least Q40  
Consensus quality: 227267 bases at least Q30  
Consensus quality: 228667 bases at least Q20  
Estimated insert size: 237036; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a working draft sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.  
1 235996: contig of 235996 bp in length.  
Location/Qualifiers  
1. 235996  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-10H14"  
1. 2153  
/note="wgs end extension  
clone end: 5p6"  
4026..7510  
/note="wgs end extension  
clone end: 5p6"  
Complement(7625..8556)  
/note="clone boundary  
clone end: 5p6  
site: EcoRI  
end\_sequence: BH305688"  
Complement(231640..232490)  
/note="clone boundary  
clone end: T7  
site: EcoRI  
end\_sequence: BH305686"

BASE COUNT 67127 a 47400 c 49017 g 66511 t 5941 others  
ORIGIN

Query Match 79.3%; Score 23.8; DB 2; Length 235996;  
Best Local Similarity 92.6%; Pred. No. 19;  
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATTGAAGCCCAAGAAATCAGCTG 27  
DB 46506 ACATAGAAAGCCCATGAAATCAGCTG 46532

RESULT 5  
AC131002 266204 bp DNA Linear HTG 08-OCT-2002  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-304K12, \*\*\* SEQUENCING IN PROGRESS  
AC131002  
AC131002.2 GI:23195369  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
SOURCE Rattus norvegicus (Norway rat)

## ORGANISM

REFERENCE  
AUTHORS

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 266204)

Muzny, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsdorfs, S., Amth, A., Angiulli, D., Anyalebech, V., Ayegbi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebreyegorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guvera, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodson, A., Hognes, M., Hollins, B., Howells, S., Hulys, S., Hume, T., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshever, L., Louisedge, H., Lozano, R.T., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nat, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweli, O., Okunolu, G., Olarnunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Ponder, A., Popovic, D., Primas, B., Fu, L., Puzos, M., Quito, J., Rachlin, B., Reeves, K., Regier, N., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.D., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smaj, D., Sneed, A., Sodergren, R., Song, X.-Z., Sorrell, R., Soza, D., Steimle, M., Strong, R., Sutton, A., Svetek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yan, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## TITLE

JOURNAL  
AUTHORS

Unpublished

## TITLE

2 (bases 1 to 266204)  
Rat Genome Sequencing Consortium.

Submitted (16-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 266204)  
Rat Genome Sequencing Consortium.

## AUTHORS

Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Sep 19, 2002 this sequence version replaced gi:22267533.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). Each contig described in the feature table below represents a scaffold in the Atlas



JOURNAL Submitted (02-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 5 (bases 1 to 86491)  
 REFERENCES  
 AUTHORS

Birken, B., Fasnath, K., Linton, L., Nusbaum, C., Lander, E., Allen, N., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Boatin, C., Boutwell, C., Brown, A., Castle, A., Cerny, J., Cooke, P., Depayre, E., Devon, K., Dewar, K., Donelan, L., Ettemad, S., Fierstein, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gardy, S., Gensheimer, S., Geratger, K., Gilman, T., Grant, G., Hagos, B., Harris, K., Horton, L., Howland, J. C., Hu, L., Jacobot, L., Kann, L., Macdonald, P., Marquis, N., McKean, P., McGurk, A., McKernan, K., Melgrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Nathanson, A., Nafie, R., Naylor, J., O'Connor, T., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Rosello, R., Roy, A., Snyam, R., Skalesky, H., Stange-Thomann, N., Stillwell, J., Stojanovic, N., Stone, C., Strickland, C., Subramanian, A., Sun, C., Torrella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W. J., Zhao, J. and Zody, M.  
 Direct Submission

TITLE Submitted (13-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Jun 27, 1998 this sequence version replaced gi:3253125. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Only the last 86.5 kilobases of this clone are being submitted. The remainder overlaps accession number AC002531 (WIGR project L216).

## FEATURES

## SOURCE

Location/Qualifiers  
 1. 86491  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="Y"  
 /map="Y"  
 /clone="HCIT\_69\_H\_8"  
 /clone\_lib="Research Genetics/Cal Tech CTRP978SK-B (plates 1-194)"  
 complement (34, 103)  
 /rpt\_family="Alus"  
 105, 390  
 /rpt\_family="MERS7\_internal"  
 392, 911  
 /rpt\_family="MERS7\_internal"  
 951, 1240  
 /rpt\_family="Alusg"  
 1242, 2524  
 /rpt\_family="MERS7\_internal"  
 complement (2526, 2824)  
 /rpt\_family="AluY"  
 2848, 3734  
 /rpt\_family="MERS41\_internal"  
 complement (4181, 4585)  
 /rpt\_family="MERS2"  
 complement (5593, 5819)  
 /rpt\_family="MERS20"  
 5830, 5871  
 /rpt\_family="AT\_rich"  
 5835, 6244  
 /standard\_name="EST"  
 /note="acc. number AA211410, match to gi 1810064, clone zq88f03.gi"  
 6407, 6611  
 /rpt\_family="MLTLE"  
 6612, 6794  
 /rpt\_family="MSTR"  
 complement (6801, 7104)  
 /rpt\_family="Alusx"  
 7107, 7359  
 /rpt\_family="MSTR"  
 complement (7886, 8182)  
 /rpt\_family="AlusC"

## STS

8147, .8524  
 /standard\_name="BST"  
 /note="acc. number AA224051, match to gi 1844592, clone zrl1c07.r1"  
 8408, .8517  
 /rpt\_family="T11MG"  
 8582, .8645  
 /rpt\_family="AT\_rich"  
 8682, .9075  
 /standard\_name="BST"  
 /note="acc. number AA207105, match to gi 1802456, clone zq81b05.gi"

## STS

8858, .8960  
 /rpt\_family="AluYb"  
 9038, .9314  
 /rpt\_family="AluDo"  
 9528, .9762  
 /rpt\_family="MERS4"  
 complement (9764, 10075)  
 /rpt\_family="MERS4A"  
 complement (10076, 10213)  
 /rpt\_family="MERS4B"  
 complement (10233, 11146)  
 /rpt\_family="MERS4\_internal"  
 10560, .10604  
 /standard\_name="BST"  
 /note="acc. number AA204659, match to gi 1802573, clone zq81b05.r1"

## STS

11147, .11432  
 /rpt\_family="AluY"  
 complement (11441, 12658)  
 /rpt\_family="MERS4\_internal"  
 11462, .11556  
 /standard\_name="BST"  
 /note="acc. number AA204659, match to gi 1802573, clone zq81b05.r1"

## STS

12659, .12966  
 /rpt\_family="AluSg"  
 complement (12967, 13508)  
 /rpt\_family="MERS4\_internal"  
 complement (13486, 13986)  
 /rpt\_family="MERS4\_internal"  
 13989, .14199  
 /rpt\_family="MERS8"  
 complement (14200, 14279)  
 /rpt\_family="MERS4\_internal"  
 complement (14305, 14587)  
 /rpt\_family="AluY"  
 14617, .14912  
 /rpt\_family="AluSg"  
 complement (14931, 15353)  
 /rpt\_family="MERS4\_internal"  
 15358, .15647  
 /rpt\_family="AluY"  
 complement (15653, 15831)  
 /rpt\_family="MERS4\_internal"  
 complement (15833, 15965)  
 /rpt\_family="MERS4\_internal"  
 complement (15972, 16220)

## STS

Query Match 77.3%, Score 32.2; DB 9; Length 86491;  
 Best Local Similarity 89.3%; Pred. No. 36;  
 Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

## QY

2 TATGAGAGCCCAAGAAAATCATGCTGAC 29  
 DB 44376 TATGAGAGCCCAAGAAAATCATGCTGAC 44403

## REPEAT 7

AC120085/c AC120085 195296 bp DNA linear HTG 12-OCT-2002  
 LOCUS Rattus norvegicus clone CH230-490B8, \*\*\* SEQUENCING IN PROGRESS

\*\*, 5 unordered pieces.

ACCESSION  
AC120085.4 GI:23908228  
VERSION  
HTG\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
1 (bases 1 to 195296)  
Allen, C., Murre, M., Metzger, M., Lee, A., Adams, C., Alder, J.,  
Mullen, C., Allen, H., Albrooks, S., Amato, A., Angiano, D.,  
Aylward, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaru, D., Barber, M., Barnstead, M., Benham, F.,  
Biswal, K., Blair, J., Blankenship, K., Blythe, P., Brown, M.,  
Bryan, N., Burch, C., Burch, P., Buttrell, K., Calderon, E.,  
Cadenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Diaper, H., Dugan-Kocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,  
Hollins, B., Howell, S., Hui, S., Hume, J., Idubrid, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Koyar, C.,  
Kowis, C., Kraft, C.L., Lebow, H., Levan, Z., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, D.,  
Lorenz, B., Louie, L., Louie, M., Lozano, R., Lu, X., Ma, J.,  
Maheshwari, M., Mahindratne, M., Mahmood, M., Malloy, C., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, B.,  
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
Mikolajewicz, A., Miner, G., Minja, B., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, K., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nasekela, O., Okunolu, G., Olariu, S., Olanrewaju, A., Pal, S.,  
Pastermak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plummer, F., Polinder, A., Popovic, D., Primus, B., Pu, L.,  
Pizarro, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reich, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
Sander, N., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,  
Steinle, M., Strong, R., Sutton, A., Swalek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, D., Tinsley, A., Trejos, Z., Uman, K.,  
Valas, R., Vera, V., Villaena, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, K., White, F.,  
Williams, G., Wilson, R., Wlezyk, R., Woodson, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhou, D., von  
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.

TITLE  
Direct Submission

JOURNAL  
Unpublished

REFERENCE  
2 (bases 1 to 195296)

AUTHORS  
Mortley, K.C.

TITLE  
Direct Submission

JOURNAL  
Submitted (03-MAY-2002)

REFERENCE  
3 (bases 1 to 195296)

AUTHORS  
Submitted (12-OCT-2002)

TITLE  
Submitted (12-OCT-2002)

JOURNAL  
Submitted (12-OCT-2002)

# COMMENT

On Oct 12, 2002 this sequence version replaced gi:21747300.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contract: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GWMG  
Center clone name: CH230-49086  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 17595 bases at least Q40  
Consensus quality: 178795 bases at least Q30  
Consensus quality: 180615 bases at least Q20  
Estimated insert size: 177042; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 5 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 184779: contig of 184779 bp in length  
\* 184780 184879: gap of unknown length  
\* 184880 188094: contig of 3215 bp in length  
\* 188095 188194: gap of unknown length  
\* 188195 191782: contig of 3588 bp in length  
\* 191783 191882: gap of unknown length  
\* 191883 193200: contig of 1318 bp in length  
\* 193201 193300: gap of unknown length  
\* 193301 195296: contig of 1996 bp in length.

## FEATURES

source  
1. 195296  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-49086"  
1. 1322  
/note="wgs end-extension  
clone end: T7"  
misc\_feature  
5152..6277  
/note="clone boundary  
clone end: T7"  
site:MoI  
end sequence: RXBRP04TV"  
complement(167606..168692)  
/note="clone boundary  
clone end: Sp6"  
site:MoI  
end sequence: RXBRP04TV"  
173865..177056  
/note="wgs end-extension  
clone end: Sp6"  
179471..180564  
/note="wgs end-extension  
clone end: Sp6"

misc\_feature 180827..182460  
/note="wgs\_end\_extension  
clone\_end:Sp6"  
misc\_feature 182511..184779  
/note="wgs\_end\_extension  
clone\_end:Sp6"  
misc\_feature 184880..186190  
/note="wgs\_end\_extension  
clone\_end:Sp6"  
misc\_feature 188195..189223  
/note="wgs\_end\_extension  
clone\_end:Sp6"  
BASE COUNT 49922 a 41149 c 42292 g 48361 t 13572 others  
ORIGIN

Query Match 77.3%; Score 23.2; DB 2; Length 195296;  
Best Local Similarity 89.3%; Pred. No. 35;  
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATGAGAGCCCAAGAAATCAGCTGACC 30  
DB 130550 AATATAGCCCAAGAAATCAGCTGACC 130523

RESULT 8  
AX695617/c 96595 bp DNA linear PAT 31-MAR-2003  
LOCUS AX695617  
DEFINITION Sequence 1244 from Patent WO03008583.  
ACCESSION AX695617  
VERSION AX695617.1 GI:29418769  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
Morris, D.W. and Engelhard, E.K.  
Novel compositions and methods for cancer  
Patent: WO 03008583-A 1244 30-JAN-2003;  
Sages Discovery (US)  
FEATURES  
source 1..96595  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 27524 a 20559 c 21158 g 26915 t 439 others  
ORIGIN

Query Match 74.0%; Score 22.2; DB 6; Length 96595;  
Best Local Similarity 88.9%; Pred. No. 96;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATGAGAGCCCAAGAAATCAGCTGACC 29  
DB 30527 ACGAGAGCCCAAGAAATCAGCTGACC 30501

RESULT 9  
AC010612 139873 bp DNA linear PRI 20-JUL-2001  
LOCUS AC010612  
DEFINITION Homo sapiens chromosome 5 clone CTB-91C24, complete sequence.  
ACCESSION AC010612  
VERSION AC010612.8 GI:14971174  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 139873)

AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 139873)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 4 (bases 1 to 139873)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (26-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Jul 20, 2001 this sequence version replaced gi:14277265.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.sngc.stanford.edu  
Quality: Phrap Quality >=40 99.6% of Sequence;  
Estimated Total Number of Errors is 0.4.  
FEATURES  
source 1..139873  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTB-91C24"  
Location/Qualifiers  
1..139873

BASE COUNT 39291 a 31188 c 29449 g 39945 t  
ORIGIN

Query Match 74.0%; Score 22.2; DB 9; Length 139873;  
Best Local Similarity 88.9%; Pred. No. 94;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATGAGAGCCCAAGAAATCAGCTGACC 29  
DB 39667 ACGAGAGCCCAAGAAATCAGCTGACC 39693

RESULT 10  
AL354862 153954 bp DNA linear PRI 16-NOV-2001  
LOCUS AL354862/c  
DEFINITION Human DNA sequence from clone RP11-8316 on chromosome 9, complete sequence.  
ACCESSION AL354862  
VERSION AL354862.11 GI:16508203  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
Gartner, P.  
Direct Submission  
Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonequest@sanger.ac.uk  
On Oct 29, 2001 this sequence version replaced gi:16501113.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em, EMBL, SW, SRA, SRA, Tr, TREMBL, Wp, WORMREP; Information on the WORMREP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormrep](http://www.sanger.ac.uk/Projects/C_elegans/wormrep). This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>. Rpl1-8316 is from the library Rpl1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>.  
VECTORS: pBAC3.6  
This sequence is the entire insert of clone Rpl1-8316. The true left end of clone Rpl1-14812 is at 6066 in this sequence. The true right end of clone Rpl1-61N16 is at 4135 in this sequence.

## FEATURES

## Location/Qualifiers

1..153954

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="9"

/clone="Rpl1-8316"

/clone\_lib="RPT-11.1"

30059..30467

/note="Sequence from overlapping clone Rpl1-148122 (AL162585) and Rpl1-61N16 (AL446024). Assembly confirmed by restriction digest."

40170..40389

/note="Sequence from overlapping clone Rpl1-148122 (AL162585) and Rpl1-61N16 (AL446024). Assembly confirmed by restriction digest."

47036..47331

/note="Sequence from overlapping clone Rpl1-148122 (AL162585). Assembly confirmed by restriction digest."

148265..148488

/note="Sequence from overlapping clone Rpl1-148122 (AL162585). Assembly confirmed by restriction digest."

44053 a 32878 c 34244 g 42779 t

BASE COUNT 44053 a 32878 c 34244 g 42779 t

ORIGIN

Query Match 74.0% Score 22.2; DB 9; Length 153954;

Best Local Similarity 88.9%; Pred. No. 93;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 71403 ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

ACAGAGCCCAAGAAATCAGCTGAC 29

## REFERENCE

3 (bases 1 to 169059)  
Bruno, D., Conn, L., Dela Rosa, M., Federpspiel, N., Foreman, P., Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R., Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J., Yu, S. and Davis, R.W.  
Submitted (08-Oct-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

## REFERENCE

4 (bases 1 to 169059)  
Bruno, D., Conn, L., Dela Rosa, M., Federpspiel, N., Foreman, P., Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R., Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J., Yu, S. and Davis, R.W.  
Submitted (20-Nov-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

## REFERENCE

On Oct 8, 1999 this sequence version replaced gi:5764710.  
USA

## REFERENCE

Location/Qualifiers

## FEATURES

1..169059

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="8"

/clone="Rpl1-4K16"

/clone\_lib="R human BAC library 11"

49086 a 33519 c 32897 g 53557 t

BASE COUNT 49086 a 33519 c 32897 g 53557 t

ORIGIN

Query Match 74.0% Score 22.2; DB 9; Length 169059;

Best Local Similarity 88.9%; Pred. No. 93;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 116058 ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

USA

3 (bases 1 to 169059)

Bruno, D., Conn, L., Dela Rosa, M., Federpspiel, N., Foreman, P., Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R., Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J., Yu, S. and Davis, R.W.

Submitted (08-Oct-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

4 (bases 1 to 169059)

Bruno, D., Conn, L., Dela Rosa, M., Federpspiel, N., Foreman, P., Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R., Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J., Yu, S. and Davis, R.W.

Submitted (20-Nov-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

On Oct 8, 1999 this sequence version replaced gi:5764710.

USA

Location/Qualifiers

1..169059

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="8"

/clone="Rpl1-4K16"

/clone\_lib="R human BAC library 11"

49086 a 33519 c 32897 g 53557 t

BASE COUNT 49086 a 33519 c 32897 g 53557 t

ORIGIN

Query Match 74.0% Score 22.2; DB 9; Length 169059;

Best Local Similarity 88.9%; Pred. No. 93;

Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 116058 ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084

ATGAAATCAGAAATCAGCTGAC 116084



TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, D., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (01-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 169060)

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (08-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 169060)

TITLE  
JOURNAL  
COMMENT

Submitted (19-DEC-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 19, 2002 this sequence version replaced g1:24796931.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Project Information  
Center project name: L12361  
Center clone name: 4\_K\_16

Location/Qualifiers  
1. 169060  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-4K16"  
/clone\_lib="RP11-11 Human Male BAC"

```

repeat_region      2..304
/rpt_family="THE1B"
repeat_region      306..750
/rpt_family="THE1B-int"
repeat_region      751..862
/rpt_family="THE1B"
complement(1171..1606)
/rpt_family="MTT1C"
repeat_region      1732..1823
/rpt_family="L2"
complement(2962..4547)
/rpt_family="L1PA3"
5084..5193
/rpt_family="L2"
complement(5329..5592)
/rpt_family="L2"
repeat_region      5593..5627
/rpt_family="L2"
complement(5628..5930)
/rpt_family="L2"
repeat_region      5966..7484
/rpt_family="L1MA8"
complement(7480..13592)
/rpt_family="L1PA6"
complement(13597..13683)
/rpt_family="L1MA8"
complement(13592..14265)
/rpt_family="L1PA2"
14266..14389
/rpt_family="L1PA2"
complement(14389..16128)
/rpt_family="L1MA8"
complement(16136..16456)
/rpt_family="MER7A"
complement(16457..16675)
/rpt_family="L1MA8"
complement(16678..17115)
/rpt_family="L1MA8"
17116..17561
/rpt_family="MSTC"
complement(17562..20016)
/rpt_family="L1MA8"
20017..20387
/rpt_family="L1MA8"
20388..21899
/rpt_family="THE1C"
21900..22208
/rpt_family="THE1C-int"
complement(22217..28367)
/rpt_family="L1PA3"
28368..28442
/rpt_family="THE1C"
complement(28443..28622)
/rpt_family="L1MA8"
complement(28647..28874)
/rpt_family="L2"
complement(28924..28991)
/rpt_family="L1R50"
complement(29529..29607)
/rpt_family="L1R41"
29608..29886
/rpt_family="MTT1F"
29950..36070
/rpt_family="L1PA6"
36188..37051
/rpt_family="L2"
complement(37165..37375)
/rpt_family="L2"
37510..37555
/rpt_family="L2"
37576..37813
/rpt_family="L1MC4a"
37808..37869

```



```

/map="17"
/clone="RP11-956N15"
/clone_lib="RBC1-11 Human Male BAC"
1..154
repeat_region
  /rpt_family="V(G) n"
  171..411
repeat_region
  /rpt_family="(CARG) n"
  complement(736..1025)
  /rpt_family="AluDb"
  1167..1190
repeat_region
  /rpt_family="AT_rich"
  complement(12668..2933)
  /rpt_family="AluSx"
  3501..3527
repeat_region
  /rpt_family="(TTTC) n"
  4314..4457
repeat_region
  /rpt_family="MER5A"
  5076..5147
repeat_region
  /rpt_family="MER116"
  6032..6119
repeat_region
  /rpt_family="MER21B"
  6194..6248
repeat_region
  /rpt_family="(TCCC) n"
  6308..6333
repeat_region
  /rpt_family="(CAAAA) n"
  6540..6625
repeat_region
  /rpt_family="L2"
  7184..7718
repeat_region
  /rpt_family="MER34"
  8026..8311
repeat_region
  /rpt_family="LTR9"
  8312..8572
repeat_region
  /rpt_family="AluY"
  8573..8918
repeat_region
  /rpt_family="LTR9"
  11276..11579
repeat_region
  /rpt_family="L1M3"
  11998..12012
repeat_region
  /rpt_family="L1M6"
  12013..12322
repeat_region
  /rpt_family="AluSx"
  12323..12521
repeat_region
  /rpt_family="L1M6"
  12523..12786
repeat_region
  /rpt_family="MER51B"
  complement(12787..12984)
  /rpt_family="L1M1"
  complement(12985..13292)
  /rpt_family="AluY"
  complement(13293..13618)
repeat_region
  /rpt_family="L1M1"
  complement(13589..14719)
  /rpt_family="L1M2"
  14757..14827
repeat_region
  /rpt_family="L1M4c"
  14828..15373
repeat_region
  /rpt_family="(PARATG) n"
  15374..15445
repeat_region
  /rpt_family="L1M4c"
  15446..15580
repeat_region
  /rpt_family="(PARATG) n"
  complement(15581..17070)
  /rpt_family="L1M2"
  complement(17578..17868)
repeat_region
  /rpt_family="L1M4c"
  complement(18015..18317)
  /rpt_family="AluSx"
  complement(18320..18520)
  /rpt_family="L1M4c"
  18564..19069
repeat_region
  /rpt_family="MTY1G1"
  20583..20603
repeat_region
  /rpt_family="AT_rich"

```

```

repeat_region 22373..22669
  /rpt_family="AluDo"
  22874..23032
repeat_region
  /rpt_family="L1M4"
  23242..23324
repeat_region
  /rpt_family="(TATG) n"
  complement(23398..23711)
  /rpt_family="AluSg"
  complement(24285..24447)
  /rpt_family="AluDo"
  24448..24478
repeat_region
  /rpt_family="(TTTA) n"
  complement(24479..24593)
  /rpt_family="AluDo"
  complement(24666..24713)
  /rpt_family="MADB1"

```

```

Query Match 74.0%; Score 22.2; DB 9; Length 181607;
Best Local Similarity 88.9%; Pred. No. 92;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 ATAGAGGCCAGAGAAATACGCTGAC 29
DB 77020 ACAGAGGCCAGAGAAATACGCTGAC 77046

```

```

RESULT 14
AE016956/c 304454 bp DNA linear BCT 28-MAR-2003
LOCUS
DEFINITION
  Enterococcus faecalis V583, section 10 of 11 of the complete
  genome.
ACCESSION
  AE016956 AE016830
VERSION
  AE016956.1 GI:29344791
KEYWORDS
  Enterococcus faecalis V583
  Enterococcus faecalis V583
  Bacteria; Firmicutes; Lactobacillales; Enterococcaceae;
  Enterococcus.
SOURCE
  1 (bases 1 to 304454)
  Paulsen, I., Banerjee, L., Myers, G.S.A., Nelson, K.E., Seshadri, R.,
  Read, T.D., Fouts, D.B., Eisen, J.A., Gill, S.R., Heidelberg, J.F.,
  Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, J., Beanan, M.,
  Daugherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R.,
  Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J.,
  Khouri, H., Uetzerback, T., Radune, D., Ketchum, K.A., Daugherty, B.A.,
  and Fraser, C.M.
  Role of Mobile DNA in the Evolution of Vancomycin-Resistant
  Enterococcus faecalis
  Science 289 (5615), 2071-2074 (2003)
  12663927

```

```

TITLE
  JOURNAL
  PUBLISHED
  REFERENCE
  AUTHORS
  2 (bases 1 to 304454)
  Paulsen, I., Banerjee, L., Myers, G.S.A., Nelson, K.E., Seshadri, R.,
  Read, T.D., Fouts, D.B., Eisen, J.A., Gill, S.R., Heidelberg, J.F.,
  Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, J., Beanan, M.,
  Daugherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R.,
  Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J.,
  Khouri, H., Uetzerback, T., Radune, D., Ketchum, K.A., Daugherty, B.A.,
  and Fraser, C.M.
  Direct Submission
  Submitted (03-FEB-2003) The Institute for Genomic Research, 9712
  Medical Center Dr, Rockville, MD 20850, USA
  Location/Qualifiers
  1..304454
  /organism="Enterococcus faecalis V583"
  /mol_type="genomic DNA"
  /strain="V583"
  /db_xref="taxon:226185"
  complement(94..2229)
  /locus_tag="EF2857"
  complement(94..2229)
  /locus_tag="EF2857"
  /note="similar to GB:M96803, GB:S65762, GB:D17086,
  SP:001082, and PID:338443; identified by sequence

```

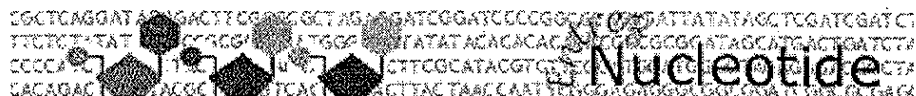
```

FEATURES
  source
  gene
  CDS

```

similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="penicillin-binding protein 2b"  
/protein\_id="AA082549.1"  
/db\_xref="GI:29344792"  
/translation="MKKSFIDKIKTKDAANSQGRKPHIPRLNLFVILFVLS  
LIVRLGYLQVEGEFERNKILTNSSLOITTPSPRGQIVDSQKVIYNSKMLATYTR  
GKNIIEGDLIPANKYENILNVPDPLNDPKDYLWLANPENLKAOSKRLTDQKED  
EKNKITDEGTITLAKAVKVEYPIEIAFDRTLOAVTIFKRNAAQSMNTVFINKNGVT  
KEGIATIGHTLAEISGVSTGTWTRDYSQSGALRSLLGVSTKOGSLPABEDYELK  
GYARNRVGTSYIEKQYEDYLOSKAKSEVVDNKGITVSQTPISKESGSLKTLID  
SNQNKVDETLQKNYSQIVKTIQIPSENAIVVAMNFGVAILLAGSGFHDLLTGVTLP  
NPLAFLNSEVGSVVAGTLLAGYETGVIKAGDYLIDELAILAGSNPASAQWMASSG  
TMOQLTAEQALERTSSNAYMKLVFKMAGVYNNMIFPYEGDDYFVKELRAFAEY  
MGTKGIDIPGETTIGIONDFOSSAPQGNLIDLSFGQYDTYSLAOLAOYVSTVAN  
NGIRVOHVAGIYGVNDENGALGKTLKLEPKYLNKVNISSEQIGLQGFYVNVGT  
SPPTTARGKSKDFTSIKATGTRFETATNAGVNTTANSNLVAIAPENPAIATSV  
LPHINDASKNQOTIAEVLAEIEMETIKY"  
/complement(2394..4331)  
/gene="thrs"  
/locus\_tag="EF2858"  
/complement(2394..4331)  
/gene="thrs"  
/locus\_tag="EF2858"  
/complement(2394..4331)  
/note="identified by match to PPM protein family HMM  
PF02824"  
/codon\_start=1  
/transl\_table=11  
/product="threonyl-tRNA synthetase"  
/protein\_id="AA082550.1"  
/db\_xref="GI:29344793"  
/translation="MSIHITPPDGAIVKPPDSGITTFEDVAKSISNSIAKAKLAGKENG  
LIDLDPIVEDSLAEIVPDHEDALGILHSSAHMLANALRLFPNIKGVGPAIDSG  
FYDNDGHSPTAEDLPALIBAMKTKENNVIVKRTISPAALRLEADDPKYEIL  
TDLPEELITVYDQCDPVDLCGVAHPSTGRIOVRLSLVAGATYRGSNDMMQRIT  
GTFPFXKDLKEPTIKREKAKERDKKGLDLFNVSQVSGSLPFLPKGATIRRT  
ISRYIVDKISLGYOHVYTPIMADVELYTSQWHDYHEDMFPMDMGSEMLVLRPM  
NCPHMMVYKNDIHSYRELPRIALGMMHREYKSGALSGLQVREMLNDHETVPR  
DOIODEFKRTLEHVAIVADFNITDIFRFLSYRDPNNTDKYFDDMAKKAQMLKAA  
MDELDELFEASGEAAFYGPKLDNOVKTLGMEHTLSTQIDPLPBPDPDLTYVEDG  
KNTHPRVVHRCIVSTMERFAYLFTVYGCAPPTMLAPLOATITTYSEAEHAYAEI  
KERLQAGIRVBYDSENERGKIRIASQOKVPYQVLVDGKEMEDATVAVRRYSKET  
SVEDSLIFDSMAAEVHANSR"  
/complement(4728..4907)  
/locus\_tag="EF2859"  
/complement(4728..4907)  
/locus\_tag="EF2859"  
/note="identified by match to PPM protein family HMM  
PF01361"  
/codon\_start=1  
/transl\_table=11  
/product="4-oxalocrotonate tautomerase, putative"  
/protein\_id="AA082551.1"  
/db\_xref="GI:29344794"  
/translation="MPYHVELLEGRTHEQLTNMYKDI TEAVSKNAGPKENIHIVN  
ELKKRYIAQSGSWKCK"  
/complement(5072..6496)  
/locus\_tag="EF2860"  
/complement(5072..6496)  
/locus\_tag="EF2860"  
/note="identified by match to PPM protein family HMM  
PF03734"  
/codon\_start=1  
/transl\_table=11  
/product="BrlK/Ybis/Ycfs/Ynhg family protein, putative"  
/protein\_id="AA082552.1"  
/db\_xref="GI:29344795"  
/translation="WNTTRRQSGKNNKKKVIITSLVGLAVAGSYYVFSHFPLPT  
KNGVSVGNVMAAEKLAQVQTEVVVQGTKEEKIQLEPKQLQOKFLKHLHS  
SKYGLPNAFKKELEAKLATSPEGEKSKVASIRRGNGTEIIVBBSGGTVVDORL  
NNOITADVAGNGOYNAKDPYKAPLTKPOCTIKATLTITNANTITTYDINGEK  
VAFKTOIONVANDGCTINKKLTITVTOLETTYSANDPVLFTDHRGTRFRKNSG

YGMIDGAKTOELVNALNSQETNATLAPLVGPKENSKIANNYEHLDDOKAYCF  
IDGKIIVTVDTIGRVRKGTAVPGEKFTIYRTDVALRQMLDGSRYSVVKKMPL  
LSQGVVTOIGIHSDSKDLKXGDKKAPKIDAGSNGCINTFTGTVSKIFVYSIDGMFV  
IITGHITDABGEFDPKVDGEEY"  
/complement(6664..7527)  
/gene="aadK"  
/locus\_tag="EF2861"  
/complement(6664..7527)  
/gene="aadK"  
/locus\_tag="EF2861"  
/complement(6664..7527)  
/locus\_tag="EF2861"  
/codon\_start=1  
/transl\_table=11  
/product="aminoglycoside 6-adenylyltransferase"  
/protein\_id="AA082553.1"  
/db\_xref="GI:29344796"  
/translation="MRTEBEPQIIMHVAQOEIRAVAGVSTNNKAPDPSQDQED  
IVTVECAEFESTATVIAKFGPLIMORKEMLTFPTERTFTFLMDFDQGRIDL  
TLCPAARKDMWHEGDSLAILLDKDEMLPPLPVASDNKATYATVDDQDFNCCBFW  
VSTVAVKLCRNELEFVATHLYEYCOQELRLLSQWAMQEPEDISYKQPKYLNKY  
TPDMDQASLIDPSSEKACNSLITKQAFEDVVAQDPANMAQPTYHLQEKKVTET  
NSRLKDLQCK"  
/complement(7649..8230)  
/locus\_tag="EF2862"  
/complement(7649..8230)  
/locus\_tag="EF2862"  
/note="similar to GP:3056802; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA082554.1"  
/db\_xref="GI:29344797"  
/translation="MWMKITSISVSEFGCLILILGLIFPGKNQPTDKPALIL  
GAVTQKSTALPSYLOARLDVATYIQHPERTPYVSGGADDPATENAMKQY  
LANKVPERKQILLENKATRKTNVNSKLSFKELVITSDFMTKSOQLAKRAGIS  
NVSGITASSFPKQIKNPGRELISGLALFDW"  
/complement(8388..9332)  
/locus\_tag="EF2863"  
/complement(8388..9332)  
/locus\_tag="EF2863"  
/note="similar to GB:X59372, GB:X15506, SP:P17482,  
SP:P28356, PID:32391, PID:32398, GB:X59372, GB:X15506,  
SP:P17482, SP:P28356, PID:32391, PID:32398;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="endo-beta-N-acetylglucosaminidase"  
/protein\_id="AA082555.1"  
/db\_xref="GI:29344798"  
/translation="MGHQLSTHPFNKKTGIISTFLSILVGLFALPAFEVOAAT  
VTPTVAVYEVNNDPNNVGVKTYLAGNOAPDGLIFANINDVTKPKYILNRY  
VQOTLINEAQIIRPVORAGTKVLLSLIGNBSGAFAPPTYESADAPAAOLEOVNNTY  
HDSGIDPDDEYAEYKNGCTGPONNSSITLQRLNDKDKLITFNIGRANNSA  
NPQMSHLIDANPFTYTNWFPQIAGHPASRLGSAVGVNOLAAQVARTABOY  
GIYIMYLPFGKDSASYISAATQELGKKNYSPLVPP"  
/complement(9611..10429)  
/locus\_tag="EF2864"  
/complement(9611..10429)  
/locus\_tag="EF2864"  
/note="similar to GP:23094472; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved domain protein"  
/protein\_id="AA082556.1"  
/db\_xref="GI:29344799"  
/translation="MLAFVLAVALSITACREPKKVTASTERSKSVZETNEKTSST  
IDKNEQASSSVSNESVKNKEPDLADNNSQLTAVADITPAINGDPTTLVGIKMKK  
GESLIHPDGSNTGMIITDSDPDESRTSLISIKFGPAGALLXIGIENDNGO  
SDTKRPLLTGDSGNYPABEYVRSSEENODONEAKKIIIDOKKATYIRKTLTDA  
ENDPTIGRPGNGNDPFPRAQSKQMAANGTGTGFRFVSQCAVATTPARKG"  
/complement(10676..11395)  
/locus\_tag="EF2866"



Entrez	PubMed	Nucleotide	Protein	Genome	Structure	PMC	Taxonomy	Books
Search	Nucleotide	<input type="text"/>	for	<input type="text"/>	Go			
		Limits	Preview/Index	History	Clipboard	Details		
Display	default	Show	20	Send to	File	Get Subsequence		

## Links

```

LOCUS       AX028511                31 bp    DNA        linear        PAT 16-SEP-2000
DEFINITION  Sequence 14 from Patent WO0042216.
ACCESSION   AX028511
VERSION     AX028511.1   GI:10189694
KEYWORDS     .
SOURCE      Homo sapiens (human)
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kusk,P.
  TITLE     Genetic predisposition
  JOURNAL   Patent: WO 0042216-A 14 20-JUL-2000;
            OSTEOMETER BIOTECH AS (DK) ; KUSK PHILIP (DK)
FEATURES             Location/Qualifiers
     source          1..31
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
1 atagtgaaaa cttgtgtaat tatgaaattt t
//

```

Disclaimer | Write to the Help Desk  
NCBI | NLM | NIH

Dec 11 2003 13:29:08



CGCTCAGGATACGACTTGGGCTAGAGGATCGGATCCCGGCTGCTGATTATATAGCTCGATCGATCT  
TTCTCTATATCTTCGATATGGGATATATACACACACACGCGCGGATAGCATGACTGATCT  
CCGCACTTTCCTTCGATATGGGATATATACACACACACGCGCGGATAGCATGACTGATCT  
CAGAGACTTTCCTTCGATATGGGATATATACACACACACGCGCGGATAGCATGACTGATCT

Nucleotide

Entrez	PubMed	Nucleotide	Protein	Genome	Structure	PMC	Taxonomy	Books
Search: Nucleotide		for				Go	Clear	
Limits		Preview/Index		History		Clipboard		Details
Display	default	Show	20	Send to	File	Get Subsequence		

1: AX028510. Sequence 13 from ...[gi:10189693]

Links

LOCUS AX028510 30 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 13 from Patent WO0042216.  
ACCESSION AX028510  
VERSION AX028510.1 GI:10189693  
KEYWORDS .  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Kusk, P.  
TITLE Genetic predisposition  
JOURNAL Patent: WO 0042216-A 13 20-JUL-2000;  
OSTEOMETER BIOTECH AS (DK) ; KUSK PHILIP (DK)  
FEATURES  
source Location/Qualifiers  
1..30  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
// 1 atatagaagc ccaagaaaaa tcagctgacc

Disclaimer | Write to the Help Desk  
NCBI | NLM | NIH

Dec 11 2003 13:29:08